SEE FOLLOWING NOTE

PLEASE NOTE THAT PAGE REFERENCES MAY APPEAR IRREGULAR; HOWEVER, TEXT IS COMPLETE AS PRINTED.

VOIR NOTE CI-APRÉS

VEUILLEZ NOTER QUE LA PAGINATION PEUT SEMBLER IRRÉGULIÈRE; CEPENDANT, LE TEXTE IMPRIMÉ EST INTÉGRAL..

micromedia
a division of IHS Canada

20 Victoria Street
Toronto, Ontario M5C 2N8
Tel.: (416) 362-5211
Toll Free: 1-800-387-2689
Fax: (416) 362-6161
Email: info@micromedia.on.ca

OPTIMIZATION OF FIELD AND LABORATORY METHODS FOR

BENTHIC INVERTEBRATE BIOMONITORING

FINAL REPORT

Barry R. Taylor

Taylor Mazier Associates

R.R. # 3, St. Andrews, N.S.

Prepared for

Canada Centre for Mineral And Energy Technology

555 Booth Street

Ottawa, Ontario

K1A 0G1

January 1997

Summary

The Aquatic Effects Technology Evaluation (AETE) program commissioned a technical evaluation of field and laboratory methods for collection and enumeration of benthic invertebrates for biological monitoring at mine sites. The objective of the technical evaluation was to critically review the recent literature on field and laboratory methods for sampling benthic invertebrates, compare various methods and approaches, and recommend the most cost-effective methods for biomonitoring of metal mines in Canada. The best methods are defined as those that return the greatest amount of sensitive, relevant, reliable data for the least cost. Sensitivity is the most important of these attributes because sensitive methods can act as early warning systems of impending ecosystem damage, and are more likely to detect subtle effects of chronic, low-level metals loadings.

Study Design

The classic spatial design for biomonitoring of point sources includes one or more control or reference sites upstream from the outfall compared against a set of sites at increasing distances downstream. An alternative design, the reference areas approach, entails comparing benthic invertebrate communities at potentially affected sites against a variety of thoroughly sampled, pristine reference sites in the same physiographic region. The reference areas approach holds promise but is not yet sufficiently developed or tested for routine application at Canadian mine sites.

Biomonitoring studies at mine sites should incorporate more than one control site wherever possible. Differences between benthic invertebrate communities at the control sites can then be used to define the magnitude of natural variation, and hence the magnitude of change at downstream sites that is indicative of significant impairment. In most situations the return in information from more than two control sites probably does not justify the additional effort.

If it is not physically possible to establish upstream control sites because the mine discharges to a headwater stream or a lake outlet, a control site can be established on a comparable nearby stream, if there is one. The alternative is to use the reference areas approach by establishing a baseline of information from many streams in the region for comparison against the study stream. Both methods entail some loss of sensitivity.

Sensitivity of biomonitoring studies can be improved if habitat variables, including water depth and velocity, substratum particle size, standing crop of algae or detritus, and the organic carbon content of sediments, are measured at each site where invertebrates are collected. These data can then be used to investigate, and possibly remove, the effects of habitat variables on the densities of benthic organisms.

Currently, the best approach for assessing the effects of multiple effluents is to sample invertebrates above and below each outfall. Differences in benthic community composition between successive sites can then be ascribed to contaminants entering between them. Toxicity tests on effluents, plume delineation studies, and tracer chemicals can also help unravel the contribution of different sources, but increase the complexity and expense of the study. Even with supplementary information, it is not always possible to determine the presence, nature and extent of all the impairment and recovery zones in a river receiving multiple effluents.

Sequential Decision Plans

Sequential decision plans are a method of biomonitoring that can sharply reduce the number of samples required to detect impairment in a biomonitoring program. In a sequential decision plan, samples are collected, sorted and analyzed sequentially just until a decision can be made to classify a site as impaired or unimpaired according to predetermined levels of risk and precision. These plans can reduce the cost of biomonitoring by 50-60%, while still allowing clear-cut decisions as to whether degradation is or is not occurring.

Decision plans can only be used if a minimum effect size is agreed upon, and the approximate sampling distribution of the variable of interest is known. They say nothing about the severity of the effect. They are restricted to testing specific, predefined hypotheses, and they use only one variable to make a decision about site classification. If these limitations can be overcome, sequential decision plans are a potentially useful idea that could lead to substantial cost savings for biomonitoring.

Rapid Assessment Procedures

Rapid assessment approaches are designed to quickly identify water quality problems associated with point-source and nonpoint-source pollution and to document long-term changes in environmental conditions within a region. These procedures reduce costs by reducing sampling intensity and using simple, qualitative measures of community composition (metrics) to compare study sites against regional reference sites. Rapid assessment procedures are not statistically based and are too insensitive for use in routine mining monitoring, but they may occasionally be useful for confirmation of severe impairment.

Sample Size and Replication

Cost efficiency of benthic invertebrate monitoring programs would be dramatically improved by using much smaller samplers and increasing the number of replicates at each site. While fewer large samples are necessary to measure population densities with any given level of precision, smaller samples can be sorted more quickly and the saved effort can then be expended on collection of more replicates, which more than compensates for the small size of individual samples. For stream sampling, devices such as the T-sampler, which sample an area of 100 cm^2, should be strongly preferred over conventional devices such as the Surber sampler, which sample an area ten times larger. The effort saved by collecting smaller samples should be devoted to increasing the number of replicates from the present level of five or less to ten or more per site.

Mesh Size

The mesh size of nets used to trap benthos in the sampling devices like the Surber sampler, or of screens used to aid sample sorting in the laboratory, is of crucial importance to the effectiveness of the sampling method. Small animals, early instars of larval insects, and especially midge larvae are severely undersampled by mesh sizes of 500 μm or larger. But to sample these organisms accurately would require extremely fine meshes that are not practical or cost-effective for biomonitoring. A mesh size of 250 μm is the best compromise between efficiency and reasonably complete retention of most macro-invertebrates, and is recommended for biomonitoring at mines. Ensuring that different investigators use the same mesh size at a given site is at least as important as the actual mesh size used.

Sampler Bias

Bias refers to systematic error in the way samples represent the nature of the population or assemblage being sampled. All sampling devices are biased to some degree. Because changes in community structure at potentially disturbed sites are always determined comparatively, relative to the control sites, some bias in the sample can be tolerated if the bias is equal at all sites and most of the benthic community, and most of the organisms sensitive to the disturbance, are included in the sample. Differences in sampler bias among sites can be minimized by careful site selection, measurement of physical habitat variables, and collection of samples at all sites by one or two trained individuals.

Sample Sorting

Four kinds of methods to sorting of animals from detritus are in general use: sieves, elutriation, dyes and flotation. Sieving samples helps with subsequent sorting by separating the sample into classes containing a uniform size of particles, both benthos and detritus. Elutriators separate organisms in a sample from debris and sediments by agitating with water or air. Selective dyes that stain organisms a bright colour improve sorting efficiency by making animals easier to see. Flotation refers to placing samples in a sugar solution in which detritus sinks to the bottom while animals rise to the top. All of these various methods work to some degree, and all have limitations. Overall, facilitation methods are valuable time savers and sharply improve the cost-efficiency of sorting benthic samples, if minimum standards of specimen recovery can be met.

Subsampling increases the imprecision of population density estimates and should be used only where necessary. A small subsample still estimates the total number of organisms in the whole sample, but the uncertainty about that estimate becomes larger as the subsample gets smaller. Accuracy of population estimates for uncommon species may be seriously compromised by the reduced size of the subsample. Subsampling will also affect the estimated number of species in the sample. If samples were smaller, as recommended earlier, subsampling would be needed far less often, and these issues would be moot.

Taxonomic Resolution

The taxonomic resolution required depends on the nature of the disturbance and the scale of the investigation. Coarse taxonomy (to genus or family for insects) is sufficient to detect strong pollution effects or changes over a large geographic area. But more detailed taxonomy is necessary to detect moderate, local impairment. More complete taxonomic identifications also reveal more ecological information that can be used to interpret the nature of the stress on the benthic community.

Identification of specimens to the lowest practical level, which equates with genus for most insects and the lowest level possible without special procedures (dissection, microscopy) or reliance on specialists for all other groups, is sufficient for biomonitoring in the mining industry. More complete taxonomy, even to species for some insects, may be warranted in follow-up studies or surveys intended to examine a special problem more closely, if the added information justifies the higher cost.

The minimum level of taxonomic resolution for biomonitoring should be specified, to encourage uniformity of practice. A reference collection of benthic invertebrates should be maintained for every mine site and should be made available to consultants or researchers when each biomonitoring study is undertaken. Voucher specimens should be deposited in the reference collection after each survey. These measures would help ensure uniform and comparable taxonomy between workers and over time.

Rare Species

Benthic invertebrate communities are composed of widely uneven numbers of component species. In most water bodies the majority of species taken in any given sample are rare, collectively contributing <2% of the total number of individuals in the sample. Density estimates for rare species are unreliable, and hence of minimal utility for detecting differences between sites. The effect of these species on results of statistical analyses is almost invariably small, yet they complicate or preclude the application of many analytical methods. The abundant species contain most of the useful information in the sample, and with the exception of predators, abundant species more accurately reflect ecological conditions at the site. Deletion of statistically rare species, those for which the estimate of mean density is too imprecise to be useful, greatly simplifies analysis without significant loss of information, and should be considered as a standard practice in benthic invertebrate biomonitoring. Deleting all species that compose <1% of total numbers from all sites combined appears to be a conservative rule that is gaining acceptance.

Sommaire

Le Programme d'évaluation des techniques de mesure d'impacts en milieu aquatique (ETIMA) a commandé une évaluation technique des méthodes utilisées sur le terrain et en laboratoire pour récolter et dénombrer les invertébrés benthiques aux fins de la surveillance des effets biologiques des effluents miniers. Cette évaluation technique comportait trois grands objectifs : 1) effectuer une analyse critique de la documentation récente sur les méthodes d'échantillonnage des invertébrés benthiques sur le terrain et en laboratoire; 2) comparer diverses méthodes et approches; 3) recommander les méthodes les plus rentables (rapport coût-efficacité le plus élevé) pour la surveillance des effets biologiques des effluents des mines de métaux au Canada. Les meilleures méthodes sont celles qui se révèlent les plus sensibles et qui fournissent au moindre coût le plus de données fiables et pertinentes. La sensibilité est la caractéristique la plus importante, car les méthodes sensibles peuvent jouer le rôle d'un système d'alerte rapide en indiquant la survenue imminente de dommages écosystémiques et sont les plus susceptibles de détecter les effets subtils des faibles charges polluantes chroniques en métaux.

Plan de l'étude

La méthode la plus couramment utilisée pour la surveillance biologique des sources ponctuelles de polluants consiste à comparer les caractéristiques d'un ou de plusieurs sites témoins (ou de référence) situés en amont des sources émettrices à celles d'une série de sites situés à des distances diverses en aval de ces mêmes sources. Une autre approche, dite de zones témoins, consiste à comparer les communautés d'invertébrées benthiques de sites potentiellement contaminés à d'autres communautés bien caractérisées de sites témoins non touchés se trouvant dans la même région géomorphologique. Cette dernière approche présente un potentiel intéressant, mais elle nécessite encore des améliorations et n'est pas encore suffisamment éprouvée pour être utilisée couramment aux fins de la surveillance biologique des effets des activités minières au Canada.

Dans la mesure du possible, les études de surveillance biologique dans les sites miniers devraient prévoir la sélection de plus d'un site témoin. Dans ces conditions, les différences relevées entre les communautés d'invertébrés benthiques des sites témoins permettent d'apprécier l'importance de la variabilité naturelle et, dès lors, l'ampleur des changements révélateurs de perturbations significatives dans les sites situés en aval. Dans la majorité des cas, il est inutile d'utiliser plus de deux sites témoins, car le gain d'information ne justifie pas l'investissement d'efforts additionnels.

S'il s'avère physiquement impossible d'établir des sites témoins en amont des sources émettrices parce que la mine rejette ses effluents dans un cours d'amont ou dans l'émissaire d'un lac, on peut choisir un site témoin sur

un cours d'eau voisin, si la chose est possible. Une autre option consiste à appliquer l'approche de zones témoins pour recueillir des données de base dans de nombreux cours d'eau situés dans la région en vue de les comparer aux informations recueillies dans le cours d'eau à l'étude. Ces deux façons de faire entraînent toutefois une perte de sensibilité.

On peut accroître la sensibilité des méthodes de surveillance biologique en mesurant les variables liées à l'habitat (profondeur de l'eau, débit, distribution granulométrique, biomasse des algues ou des détritus et teneur en carbone organique des sédiments) dans chacun des sites où des invertébrés sont récoltés. On peut utiliser ces données pour déterminer et éliminer, si possible, les effets des variables liées à l'habitat sur la densité des communautés d'organismes benthiques.

À l'heure actuelle, l'échantillonnage des invertébrés en amont et en aval de chaque source émettrice constitue la meilleure approche pour évaluer les effets d'effluents multiples. Les différences liées à la composition des communautés d'organismes benthiques relevées entre plusieurs sites successifs peuvent être associées aux contaminants qui sont rejetés entre chaque site. L'évaluation de la toxicité des effluents, l'étude de la dispersion des panaches et l'utilisation de traceurs chimiques peuvent aider à évaluer l'apport de différentes sources, mais elles contribuent à accroître la complexité et les coûts de l'étude. Même en disposant de données supplémentaires, il n'est pas toujours possible de détecter toutes les zones de contamination et de rétablissement dans une rivière recevant des effluents multiples et d'en préciser la nature et l'importance.

Plans de décision séquentielle

Le recours à des plans de décision séquentielle permet de réduire de façon significative le nombre d'échantillons requis aux fins de la détection d'une contamination dans le cadre d'un programme de surveillance biologique. Selon cette approche, des échantillons sont prélevés, triés et analysés de façon séquentielle jusqu'à ce qu'on soit en mesure de déterminer qu'un site donné est contaminé ou non d'après des seuils préétablis de risque et de précision. Ces plans peuvent réduire de 50 à 60 % les coûts de la surveillance biologique tout en permettant la prise de décisions claires quant à la présence ou à l'absence de contamination.

Ces plans représentent une option seulement si l'on a convenu d'un effet minimal lié à la taille et si l'on connaît la distribution d'échantillonnage approximative de la variable à l'étude. En revanche, ils ne livrent aucune indication sur l'ampleur de l'effet et permettent uniquement de vérifier des hypothèses précises préétablies. En outre, la classification des sites (contaminés ou non contaminés) est fondée sur l'analyse d'une seule variable. Si ces obstacles peuvent être surmontés, les plans de décisions séquentielle peuvent fournir des informations

fort utiles tout en réduisant de façon substantielle les coûts de la surveillance biologique.

Méthodes d'évaluation rapide

Les méthodes d'évaluation rapide permettent de détecter rapidement les problèmes de qualité de l'eau associés à des sources ponctuelles ou diffuses et de suivre les changements à long terme des conditions environnementales dans une région donnée. Ces méthodes permettent d'abaisser les coûts de la surveillance biologique, car elles prévoient une réduction de l'effort d'échantillonnage et l'utilisation de mesures qualitatives simples de la composition des communautés pour comparer les sites à l'étude à des sites témoins établis dans la même région. Les méthodes d'évaluation rapide n'ont aucune assise statistique et ne sont pas suffisamment sensibles pour être utilisées couramment dans le cadre d'un programme de surveillance des effets de l'activité minière. En revanche, elles peuvent à l'occasion se révéler utiles pour confirmer la présence d'une contamination grave.

Taille des échantillons et nombre de répétitions

On peut accroître considérablement la rentabilité (rapport coût-efficacité) des programmes de surveillance des invertébrés benthiques en réduisant la taille des échantillons et en augmentant le nombre de répétitions dans chaque site. S'il est vrai qu'il faut moins d'échantillons de grande taille pour mesurer les densités de population à quelque seuil de précision que ce soit, les petits échantillons peuvent en revanche être triés plus rapidement. Les économies de temps et d'argent ainsi réalisées permettent d'accroître le nombre d'échantillons répétés, ce qui compense largement la faible taille de chaque échantillon. Pour l'échantillonnage des cours d'eau, les échantillonneurs en T, qui couvrent une superficie de 100 cm^2, conviennent nettement mieux que les dispositifs couramment utilisés tels que les filets Surber, qui permettent d'échantillonner une superficie dix fois plus grandes. Les économies d'effort réalisées en prélevant des échantillons de plus petits devraient permettre de porter dans chaque site le nombre d'échantillons répétés de cinq ou moins (niveau actuel) à dix ou plus.

Taille des mailles

L'efficacité de la méthode d'échantillonnage est intimement liée à la taille des mailles des dispositifs utilisés pour récolter le benthos, qu'il s'agisse de filets Surber ou de tamis permettant de trier les organismes en laboratoire. Les petits organismes, les premiers stades larvaires d'insectes aquatiques et, en particulier, les larves de chironomidés, sont gravement sous-représentés dans les échantillons lorsque les prélèvements sont effectués à l'aide de dispositifs munis de mailles d'au moins 500 µm. Pour obtenir des données fiables sur ces

organismes, il faut utiliser des mailles extrêmement fines, ce qui se révèle peu pratique sur le terrain et trop onéreux pour le seuil d'efficacité visé. L'utilisation de dispositifs munis de mailles de 250 µm représente le meilleur compromis entre un seuil d'efficacité convenable et un niveau acceptable de rétention de la majorité des macro-invertébrés benthiques. C'est ce type de dispositif qui est recommandé pour la surveillance des effets biologiques des effluents miniers dans les écosystèmes aquatiques. Il est en outre au moins aussi important de veiller à ce que tous les chercheurs utilisent la même taille de maille que de connaître la taille des mailles utilisées.

Biais d'échantillonnage

On appelle biais d'échantillonnage l'écart systématique qui existe entre la façon dont les échantillons reflètent la nature d'une population ou d'une association d'organismes et la nature réelle de cette même population ou association. Tous les dispositifs entraînent un certain biais. Comme les changements de structure des communautés d'invertébrés dans les sites potentiellement contaminés sont toujours déterminés de façon relative, par comparaison avec le changements observés dans des sites témoins, on peut tolérer un certain biais si ce dernier est constant d'un site à l'autre et si la majorité des organismes qui composent la communauté benthique et la majorité des organismes sensibles à la perturbation sont inclus dans l'échantillon. Il est également possible de réduire les écarts de biais d'échantillonnage d'un site à l'autre en choisissant soigneusement les sites, en mesurant les variables physiques des habitats étudiés et en veillant à ce que l'échantillonnage dans tous les sites soit effectué par un ou deux échantillonneurs expérimentés.

Tri des échantillons

Quatre méthodes sont couramment utilisées pour séparer les organismes des détritus dans les échantillons : le tamisage, l'élutriation, la coloration et la flottation. Le tamisage facilite le tri des échantillons du fait qu'il permet de séparer leur contenu (benthos et détritus) en classes de taille uniforme. Les élutriateurs séparent les organismes des débris et des sédiments par agitation dans l'eau ou dans l'air. Les colorants sélectifs confèrent une teinte vive aux organismes capturés et les rendent plus visibles et plus faciles à séparer. La flottation consiste à placer les échantillons dans une solution de sucre; les détritus calent au fond, tandis que les organismes remontent à la surface. Toutes ces méthodes sont efficaces à des degrés divers, mais chacune comporte des lacunes. Globalement, ces méthodes permettent d'économiser beaucoup de temps et d'accroître considérablement le rapport coût-efficacité des méthodes de tri des organismes benthiques lorsque les normes minimales en matière de récupération d'organismes peuvent être atteintes.

Le sous-échantillonnage contribue à accroître l'imprécision des estimations de la densité des populations et ne devrait être utilisé qu'en dernier recours. Un petit sous-échantillon permet d'estimer le nombre total d'organismes dans tout l'échantillon, mais l'incertitude dont est entachée cette estimation croît de façon inversement proportionnelle à la taille du sous-échantillon. Pour les espèces peu communes, la réduction de la taille des sous-échantillons peut compromettre sérieusement la justesse des estimations. Le fractionnement des échantillons en sous-échantillons influe également sur l'estimation du nombre d'espèces dans l'échantillon. Le prélèvement d'échantillons de plus petite taille permet de réduire le recours au sous-échantillonnage et, dès lors, tous les problèmes énoncés ci-haut.

Seuil de précision de l'identification taxinomique

Le degré de précision taxinomique dépend de la nature de la perturbation observée et de la portée de l'étude. Une identification grossière (genre ou famille pour les insectes) suffit lorsque le but visé est de détecter des effets ou des changements importants causés par des polluants à l'échelle d'un vaste territoire. L'identification doit cependant être plus fine si l'objet de l'étude consiste à faire ressortir les effets modérés d'une contamination locale. Un niveau d'identification plus poussé permet également d'obtenir un plus grand de données écologiques, lesquelles peuvent être utilisées pour interpréter la nature du stress subi par la communauté d'organismes benthiques.

L'identification des spécimens jusqu'au niveau taxinomique le plus faible, ce qui correspond au genre pour la majorité des insectes et au plus bas niveau qu'il est possible d'atteindre sans avoir recours à des méthodes spécialisées (dissection, microscopie) ou encore à un spécialiste pour tous les autres groupes, convient habituellement pour la surveillance des effets biologiques des effluents miniers. Une identification plus poussée, pouvant aller jusqu'à l'espèce pour certains insectes, peut se révéler nécessaire lorsque les études de suivi ou les inventaires ont pour objet d'examiner plus en détail un problème spécial, si le gain d'information justifie l'augmentation des coûts.

Il convient de spécifier le seuil minimal d'identification taxinomique requis aux fins de la surveillance biologique, de façon à encourager tous les chercheurs à utiliser les mêmes seuils. Une collection de référence des organismes benthiques récoltés dans chaque site minier devrait être mise à la disposition des experts-conseils ou des chercheurs au début de chaque projet de surveillance biologique, et des spécimens en double devraient être déposés dans la collection de référence après chaque relevé. Ces mesures permettraient de maintenir un certain niveau d'uniformité taxinomique entre les chercheurs et d'un projet à l'autre.

Espèces rares

Les espèces qui composent les communautés d'invertébrés benthiques sont représentées par un nombre très variables de spécimens. Dans la majorité des plans d'eau, la majorité des espèces trouvées dans un échantillon donné sont rares, leur contribution collective au nombre total d'organismes prélevés s'élevant à moins de 2 %. Les estimations de la densité des effectifs des espèces rares ne sont pas fiables et sont dès lors peu utiles pour faire ressortir des différences entre les sites. Bien que ces espèces aient presque immanquablement peu d'effet sur les résultats des analyses statistiques, elles compliquent la réalisation de nombreuses analyses statistiques ou en interdisent complètement le déroulement. Ce sont les espèces abondantes qui renferment la majeure partie de l'information utile dans l'échantillon et, à l'exception des prédateurs, ce sont elles qui reflètent plus fidèlement les conditions écologiques dans le site étudié. L'élimination des espèces considérées statistiquement comme rares et de celles pour lesquelles l'estimation de la densité moyenne est trop imprécise pour être utile simplifie considérablement l'étape des analyses statistiques sans causer une perte d'information significative. Cette façon de faire devrait être considérée comme la norme aux fins de la surveillance biologique des invertébrés benthiques. L'élimination de toutes les espèces qui représentent moins de 1 % du nombre total d'organismes récoltés dans tous les sites semble une règle conservatrice de plus en plus reconnue parmi la communauté scientifique concernée.

TABLE OF CONTENTS

Page

1. Introduction 1
   1.1 Background and Approach 3

2. Study Design 6
   2.1 Site Selection 7
      2.1.1 Single Point Sources 7
      2.1.2 Control Sites 11
      2.1.3 Habitat Variables 14
      2.1.4 Multiple Contaminant Sources 16
   2.2 Sequential Decision Plans 18
      2.2.1 Overview 18
      2.2.2 Designing A Plan 21
      2.2.3 Advantages and Disadvantages 24
   2.3 Rapid Assessment Approaches 28

3. Field Methods 32
   3.1 Sample Size and Replication 32
      3.1.1 Overview 32
      3.1.2 Spatial Distribution of Benthic Invertebrates 32
      3.1.3 Implications of Aggregation for Sample Size 34
      3.1.4 Minimum Sample Size 40
      3.1.5 Persistence of Large Samples 41
      3.1.6 Optimal Sample Sizes for Site Comparisons 42
   3.2 Mesh Size 43
      3.2.1 Effect on Sampling Efficiency 43
      3.2.2 Persistence of Mixed Mesh Sizes 48
   3.3 Sampler Bias 50

4. Laboratory Methods 54
   4.1 Sample Sorting 54
      4.1.1 Facilitation 54
      4.1.2 Subsampling 59
   4.2 Taxonomic Resolution 62
      4.2.1 Value of Species-Level Identifications 63
      4.2.2 Statistical Considerations 71
      4.2.3 Limitations of Taxonomy 73
      4.2.4 Mixed Taxonomy 75
      4.2.5 Reference Collections 77
   4.3 Rare Species 77

5. Conclusions and Recommendations 84
   5.1 General Conclusions 84
   5.2 Recommendations 85

6. References 89

LIST OF TABLES

Table 1 Example of a sequential decision plan in tabular form. (21)

Table 2 Number of replicates needed for various sampler sizes and macrobenthos densities in order that the standard errors of replicate samples average 20% of the mean density. (36)

Table 3 Area of sediment (cm^2) that must be sorted to obtain a standard error of replicate samples averaging 20% of the mean density, as a proportion of the area that must be sorted to reach the same precision using a sampler of 1000 cm^2. (37)

Table 4 Ratios of mesh sizes, net sizes and density of drifting invertebrates captured in drift nets. (45)

Table 5 Numbers of larval chironomids from Lake Michigan sediment samples retained on screens of different mesh size. (47)

Table 6 Mesh size of capture nets or sieves used in the field to sample benthic invertebrates, as reported in the published literature. (49)

Table 7 Effect of facilitation on time required to sort various benthic invertebrate samples, based on responses to a questionnaire. (59)

Table 8 Chironomid species list from an Ohio stream (Elam's run) contaminated with mixed heavy metals. (65)

LIST OF FIGURES

Figure 1 Example of a sequential decision plan. (20)

Figure 2 Example of a sequential decision plan to detect both moderate and severe changes in water quality. (27)

Figure 3 Number of *Cheumatopsyche pettiti* larvae in replicate Surber samples taken side by side in Rock Creek, Indiana. (34)

Figure 4 Effect of desired precision (standard error as a proportion of the mean), mean density and sampler size on sampling effort. (39)

Figure 5 Relationship between total number of individuals counted in one or more subsamples and the associated percentage error at a 95% confidence level. (61)

Figure 6 Distribution of Nearctic genera in the caddisfly family Polycentropodidae along a continuum from headwater streams to large rivers, and in lentic habitats and rocky lake shores. (70)

Figure 7 Hypothetical example of a log-normal distribution of individuals among different species in a benthic invertebrate community. (78)

Figure 8 Examples of species abundances in benthic invertebrate samples. (79)

APPENDIX

Examples of taxonomic lists from benthic invertebrate samples

Acknowledgements

This report was written for the Aquatic Effects Technology Evaluation (AETE) program, a co-operative program among the Canadian mining industry and responsible agencies of the federal government and most provincial governments. The project was directed by Lise Trudel of the Canada Centre for Mineral and Energy Technology (CANMET). Thorough and helpful reviews of earlier drafts by Roy Parker (Environment Canada, Dartmouth, N.S.), William Duncan (Cominco Ltd., Trail, B.C.), Glen Watson, (Inco Ltd., Copper Cliff, Ont.), Bill Keller (Ministry of Northern Development and Mines, Sudbury, Ont.), Dennis Farara (Beak International Inc., Brampton, Ont.), Danuta Zaranko (Zaranko Environmental Assessment Services, Guelph, Ont.) and Robert Bailey (Dept. of Zoology, University of Western Ontario, London) considerably improved the final document. The author is grateful as well to Patrick Stewart (Envirosphere Consultants Ltd., Windsor, N.S.), Zsolt Kovats (Golder Associates Ltd., Calgary, Alberta) and Richard Pope (Tarandus Associates Ltd., Brampton, Ont.) for freely sharing their time and expertise on the subject of benthic invertebrate sampling. Valuable comments and direction were also received from Derek Riehm (Teck Corporation, Vancouver, B.C.), Roger Green (Dept. of Zoology, University of Western Ontario) and the other members of the AETE Technical Committee.

1. Introduction

The Aquatic Effects Technology Evaluation (AETE) program was established to review appropriate technologies for assessing the effects of mine effluents on aquatic ecosystems. AETE is a co-operative program among the Canadian mining industry, several federal government departments and eight provincial governments; it is co-ordinated by the Canadian Centre for Mineral and Energy Technology (CANMET). The program has two stated objectives: to help the Canadian mining industry meet its obligations for environmental effects monitoring in the most cost-efficient manner; and to evaluate new and established monitoring technologies that could be used for assessment of environmental effects of mining. The program is designed to be of direct benefit both to the industry and to government by evaluating and identifying cost-effective technologies to meet environmental monitoring requirements. The program includes three main areas: acute and sublethal toxicity testing, biological monitoring in receiving waters, and water and sediment monitoring.

The AETE program includes field evaluations of biological monitoring technologies to be used by the mining industry and regulatory agencies to assess the effects of mine effluents on aquatic ecosystems. The goal of the program is to recommend specific methods, or groups of methods, that will permit accurate characterization of environmental effects in the receiving waters in as cost-effective a manner as possible. A pilot field test was conducted in 1995 to fine-tune the study approach. In 1996, preliminary surveys will be carried out at seven mine sites across Canada. The field evaluation of selected monitoring methods will then take place at five of these mine sites in 1997.

Community structure of benthic macro-invertebrates, the insects, worms, molluscs and other organisms living on the bottoms of rivers and lakes, is included in the field study as an indicator of environmental quality and mine effluent effects. Field and laboratory methods for collection and enumeration of benthic invertebrates are clearly central to any biological monitoring program. Given the importance of sampling methods for the success of biomonitoring, and the variety of methods and variants presently available, the Technical Committee decided that a technical evaluation of field and laboratory methods in benthic invertebrate sampling should be carried out prior to the preparation of the field study design. The overall objective of the technical evaluation is to critically review the recent literature on methods for sampling benthic invertebrates, compare various methods and approaches, and recommend the most cost-effective methods for biomonitoring of metal mines in Canada.

This report covers elements of study design such as the choice and location of sampling sites and sampling gear, and laboratory procedures such as how invertebrates are sorted and enumerated. The

larger subject of statistical methods is not dealt with here; however, field and laboratory procedures cannot be considered in isolation from the methods that will be used to analyze the data afterward. Hence, some consideration of statistical methods that are routinely applied in biomonitoring is inevitable.

The objectives of this report are very specific. It is not intended as a how-to manual for benthic invertebrate studies, but rather as a review of some options for improving the cost-efficiency and effectiveness of biomonitoring. The goal is to point out new alternatives to conventional approaches and look for ways that biomonitoring might be optimized. Identification of data gaps where more research is needed, and recommendations for field testing are secondary objectives. From the perspective of industrial biomonitoring, the best sampling and analytical methods are those that return the greatest amount of useful, reliable data for evaluating environmental conditions in fresh waters and the most sensitive, reliable indicators of spatial or temporal changes or trends in environmental conditions, for the least cost.

Field sampling is always a trade-off between precision and cost. Efficient sampling is that which finds the best precision for the least cost. Often, by doing sampling a different way, by apportioning a fixed amount of effort differently, the precision of benthic population estimates can be improved with little or no increase in cost and effort. This principle applies to study design, sample size, sample number, mesh size, taxonomy and laboratory procedures. The idea of finding the best compromise between precision and cost is used as a guide against which different methods are evaluated in this report.

The discussion is focused on the kind and magnitude of environmental effects to be expected from wastewater discharges from metals mines in Canada, and therefore assumes the potential effects are mainly those arising from suspended sediments and heavy metals, rather than organic pollution or nutrient enrichment. The text is unavoidably weighted toward monitoring of flowing waters, in particular small to medium-sized streams and rivers, because that is where the majority of waste effluents have traditionally been discharged (with marked exceptions like the Laurentian Great Lakes) and hence where the majority of research has been done. Information from large rivers, lakes and the oceans is included where it is relevant.

The literature review was limited to works published since 1980. Particular attention has been devoted to studies from the last ten years, because many big strides have been made in that period that have potential to substantially change biomonitoring approaches. A comprehensive review of biomonitoring with benthic invertebrates, covering most of the topics in this report, was published by Rosenberg and Resh (1993). Their work provides an expert synthesis of the literature up to about 1990. Relying on the work of Rosenberg and Resh to cover the older literature, the effort in this work was concentrated on (a) updating the literature review

to include accounts from 1990 through 1996, and (b) re-examining the very general evaluation of Rosenberg and Resh (1993) in the narrower context of methods for the mining industry.

In applied sciences such as biomonitoring there is always a gap between published methods and approaches, and the methods that are actually used from day to day by biologists carrying out biomonitoring for the mining companies or regulatory agencies. Therefore, ideas from the literature survey were supplemented by discussions with consultants involved in biomonitoring. Several recent surveys of freshwater biologists and limnologists (Resh and McElravy 1993, Winterbourn 1985, Resh et al. 1985) also provide information on which methods are routinely employed in benthos sampling. Common practice is seldom a good indicator of best practice; nevertheless, it is useful to know which methods in the literature have been adopted by practising scientists and consultants.

1.1 Background and Approach

Before any detailed discussion of study design or laboratory and field methods can begin, it is necessary to look at the specific goals of biomonitoring at mine sites, and see how those goals affect the general approach taken, and hence the sensitivity and efficiency of the program. Biomonitoring practised for regulatory or environmental protection purposes at mines is intended only to assess the effects of effluents or run-off on the local aquatic environment. It is not intended as a comprehensive survey of the aquatic community nor as a broader water quality survey. The quintessential example of this kind of monitoring is routine, usually annual or semi-annual surveillance of a water course receiving wastewaters from a mine site at a single, well-defined effluent outfall. At most mines the true situation is far more complicated than that, but the fundamental ideas are the same. These annual assessments have three related objectives:

(1) to determine whether the effluent outfall is having a detrimental effect on the benthic invertebrate community downstream;

(2) to measure the severity of any detrimental effects; and

(3) to determine how far downstream the effects extend.

Detrimental effects are most often defined operationally as any significant variance from the community structure at a comparable control site or sites, usually upstream (Underwood 1991, DFO & Environment Canada 1995). Published studies recommending a particular method for benthic invertebrate monitoring must be considered in the specific context of their utility for biomonitoring at mine sites (Camacho and Vascotto 1991).

The three related objectives of a routine monitoring program lead to a set of key assumptions about the nature of the data collected and how it will be analysed. Those assumptions, in turn, have important implications for the way that sampling and sorting should be carried out to maximize sensitivity and minimize costs. Therefore, in this report, the following five assumptions were used as the basis for comparing and evaluating alternative approaches to benthic invertebrate monitoring:

(1) It is only environmental conditions within the system being sampled that are of interest. The question is very specific: What is the effect of *this* effluent on *this* stream? (Resh and McElravy 1993). There is no attempt to compare the sampled water course with others in the region (but see Section 2.1.1), and no attempt to answer general questions about the nature of effluents on streams (Stewart-Oaten et al. 1986).

(2) The presence or population densities of invertebrate species are only of interest insofar as they contribute to detecting differences between upstream and downstream sites. Sampling is not necessarily intended to provide a comprehensive inventory of the entire benthic community.

(3) The data analysis is always comparative. It is the *change* in population density or the *difference* in community structure at downstream sites compared with control sites that is of interest, not their absolute values (Stewart-Oaten et al. 1986, Ferraro et al. 1989).

(4) In the absence of spills or accidents, effluent effects observed downstream are likely to be small. Chronic, low-dose exposures to contaminants are more important than acute, high-dose exposures.

(5) The analysis is based on one set of samples, all collected more or less simultaneously, and comparisons through time are not the first priority of the program.

These assumptions stress the need for a program that is sensitive, i.e., that can reliably detect relatively small changes in the benthic community. Biological responses to strong disturbances are relatively easy to detect. It hardly matters whether sampling and analytical methods are sensitive when the impairment of the benthic community is large and conspicuous. Any method, even if statistically suspect, inefficient and suboptimal, will suffice for that purpose.

But there is only limited value in biomonitoring to document the effects of severe disturbance, which is obvious in any case. It is far more useful to use biomonitoring as an early warning system, to signal ominous changes in environmental quality at their inception, so that corrective action may be taken before major environmental damage occurs (Bunn 1995, Humphrey et al. 1995). This latter use of biomonitoring is the intent of routine

surveys at mines and industrial sites where the quality of aquatic ecosystems is monitored as a normal part of environmental vigilance. Traditional techniques developed for organic enrichment, or extended from them, may be less suitable to monitoring the smaller and more subtle effects to be expected from low-level metals contamination or other disruptions associated with mining.

A major thesis explored in this report is that sensitivity and cost-effectiveness of biomonitoring may be limited as much by inertia as by scientific understanding. Since conventional methods for assessment of pollution with benthic invertebrates became established, researchers have turned their attention to the particular goals of biomonitoring, and many new developments have been proffered to improve the sensitivity, efficiency and economy of biomonitoring with macro-invertebrates. Yet there has been reluctance on the part of field workers to embrace these new methods, apparently because of a kind of methodological tradition, perhaps coupled with the influence of training in conventional science. Substantial departures from methods used for scientific studies of stream ecology may be necessary if biomonitoring at mine sites is to be optimized.

2. Study Design

When biomonitoring is used to assess effects of mines or other industries, the study design will generally be one of three broad types, depending on the objectives and hypotheses of the study (Environment Canada 1993):

(1) Spatial Design. This design involves a comparison between benthic communities at sites potentially affected by an effluent or other source of pollution or disturbance, and reference or control sites not so affected. Also known by the semantically unfavourable name of Control-Impact design, this is the most commonly used design for routine monitoring at mine sites.

(2) Temporal Design. This design involves comparison of benthos communities over time, such as before and after start-up of a new industry, or before and after some change in operation, such as an increase in effluent volume or an improvement in wastewater treatment. Hence the common name Before-After design. Long-term benthos monitoring to check for gradual improvement or deterioration of water quality would also be included in this class.

(3) Site by Time Design. More commonly known by the acronym BACI (Before-After-Control-Impact), this design is a combination of the preceding two. In a classic BACI design (Green 1979), one or more potentially affected sites and control sites are sampled before and after a new source of disturbance begins, and the change in the difference between sites is the comparison of interest. BACI designs are statistically challenging and have engendered a great deal of debate and numerous variations involving multiple sampling sites, repeated measures, and fixed, simultaneous, or random sampling times (Underwood 1991, 1992, 1993, Faith et al. 1991, 1995, Stewart-Oaten et al. 1986, 1992, Smith et al. 1993).

The simple spatial design, or a more complex version of it, is probably the most widely used for biomonitoring of industrial effluents and similar local disturbances. The consensus among practitioners of biomonitoring in Canada is that this design is satisfactory for most benthic invertebrate studies examining effects of point sources if appropriate modifications are made regarding the number and location of sampling sites (Environment Canada 1993). The details of advanced BACI designs and derivatives and the often intense disagreements over the statistically superior design are not covered here. These complex designs are probably most useful for situations such as environmental impact assessments of large projects, where time and resources are sufficient and a specific effect is being investigated. However, some of the issues concerning BACI designs apply equally to other study designs and are mentioned in the ensuing discussion where relevant.

2.1 Site Selection

Sampling design is perhaps the most widely discussed element of benthic invertebrate sampling, and most practitioners appear to be aware of the importance of careful site positioning. Therefore, no comprehensive critique is necessary here. Rather, attention is devoted to a few key points that could help improve the efficiency of biomonitoring.

2.1.1 Single Point Sources

The importance of choosing the right number and location of sampling sites for benthic invertebrate monitoring is widely recognized, and the general protocol, at least for simple point-source effluents on lotic systems, is well established (Anderson 1990, Klemm et al. 1990). The key elements of this plan are based on advice in Cairns and Dickson (1971):

(1) Always have at least one (preferably two) reference or control sites located outside the zone of influence of all effluents for comparison with points below the discharge. This site should be directly above the effluent outfall in streams and rivers or just outside the zone of influence of the effluent in lakes. It is advisable to have a second control site well upstream or far away from the first, to provide a baseline of spatial variation in community structure.

(2) Establish a site directly below the pollution source. Ensure that the site is within the plume of the discharge.

(3) Establish more sites at increasing distances downstream in rivers or away from the discharge point in lakes. It is best if these sites are spaced at approximately exponentially increasing distances to establish the spatial extent of effluent effects and the length of the recovery zone.

(4) Ensure that all sampling sites are as ecologically similar as possible to maximize comparability of results. Substratum particle size, current velocity (or wave action) and depth are the most important habitat features. Sampling locations should also be located where benthos is not affected by atypical conditions, such as those created by bridge crossings or dams, unless effects of these conditions are part of the study design.

(5) Sample benthic invertebrates as closely as possible to sites where chemical or physical measurements are taken.

(6) Attempt to locate sampling sites where they can all be sampled at approximately the same time, preferably the same day.

The sampling design for environmental effects monitoring at pulp and paper mills (DFO and Environment Canada 1993) is a variant on the classic design, in which three areas are defined for sampling: a reference area that is similar to the exposed monitoring sites but that is not exposed to the effluent; a near-field exposure area beyond the immediate region of discharge but still within a high concentration of the effluent; and a far-field area exposed to much lower effluent concentrations than the near-field sites but still within the 1% effluent dilution boundary. The environmental effects monitoring program differs from the traditional approach in that comparisons among areas, each of which may contain more than one sampling station, is considered more important than comparisons among individual stations. The analysis of variance therefore uses variation between stations as the error term, rather than variation among replicates, which is just sampling error. This design is more labour intensive than the traditional approach and does not facilitate defining the gradient of pollution effects.

In contrast to these variations on the classic approach to biomonitoring studies, the *reference areas* approach represents an entirely different paradigm for environmental monitoring. Instead of one or several control sites or zones for each study, this approach entails comparing benthic invertebrate communities at potentially affected sites against a variety of thoroughly sampled reference sites in the same physiographic region. The benthic invertebrate communities at the reference sites are taken to represent the normal condition, unaffected by human influence, and the nature and degree of impairment at affected sites is determined by how far their benthic communities depart from the those at reference sites. The reference areas approach was developed in conjunction with rapid assessment procedures (Section 2.3), but can be used with quantitative sampling methods as well.

At its simplest, the reference areas approach replaces the usual control sites with regional reference sites. This system typically includes information from a large number of reference sites within a particular region, which is compiled into a uniform data base; such a system is presently in use in the United States to facilitate quick environmental quality surveys embracing many streams and rivers (Plafkin et al. 1989). The reference sites thereby define not only the normal invertebrate community, but also the range of variation to be expected in the absence of human intervention, which helps to assess the severity of the impairment at the site being compared.

In the U.K., a more advanced system known as RIVPACS (River Invertebrate Prediction and Classification System) has been developed (Furse et al. 1984, Wright et al. 1984, 1988, Moss et al. 1987). RIVPACS is based on simultaneous sampling of benthic invertebrates and physical and chemical characteristics at riverine sites across Great Britain. The resulting data base now contains data from 438 sites on 80 rivers, with measurements of up to 28 physical-chemical variables at each site (Cash 1995). At least half these sites have been sampled in three different seasons, with qualitative and quantitative methods, and species-level identifications wherever possible (Furse et al. 1984). Subsets of the main data set, using family-level identifications, qualitative abundance categories and a few key environmental variables, are used for most analyses (e.g., Armitage et al. 1987).

Multivariate analysis is used to classify the sites and identify key species that are most useful for classification. Multiple discriminant analysis is then used to produce discriminant functions based on the environmental variables that best separate site groupings identified in the biological ordination. The resulting models can be used to predict the composition of benthic invertebrate communities at other sites where only physical-chemical variables have been measured.

The strength of this approach is that it produces a robust and powerful indicator of expected community structure that controls for the confounding effects of environmental variation among sites. Hence, the method could be used to predict the fauna at a site receiving mine wastewaters, and the difference between the predicted and observed community structures would indicate the degree of impairment of the community. The predicted community can also be seen as a "target" or goal for a remediation program or wastewater treatment upgrade. Changes in the observed community toward the target community structure provide a quantitative indicator of improving environmental quality (Cash 1995).

A further extension of the reference sites approach is under development in the Laurentian Great Lakes. The Benthic Assessment of Sediment (BEAST) uses an approach similar to that used in RIVPACS, but with important additional procedures involving sediment toxicity testing (Reynoldson et al. 1995). In this method, patterns in lake macroinvertebrate community structure at nearshore sites are explored using ordination and cluster analysis. The ordination scores are then correlated with measured environmental variables to determine which sediment and water quality characteristics are most strongly associated with patterns in macro-invertebrate community structure. Finally, the site groupings from the classification are related to the key environmental variables using multiple discriminant analysis. As in RIVPACS, the results of the discriminant analysis support a model that can be used to predict community structure at other, potentially contaminated sites for which environmental data are available (Cash 1995).

The reference areas approach is a radical departure from the standard method employing only a few upstream control sites. The predictive capability of the method is attractive, but there are a number of considerations that constrain its utility for biomonitoring at Canadian mine sites. Chief among them is the need to amass a large, consistent data base on benthic invertebrates and chemical-physical variables at many sites. No such data bases are presently available, although Cash (1995) recommended the reference areas approach for monitoring environmental quality in the Northern Rivers Basin Study. The creation of such a data base would entail a considerable and probably prohibitive cost for an area of any substantial size. Further, the models are not reliable for environmental conditions outside the range on which they are founded. For a geographic area as large and diverse as Canada, this limitation means that a separate data base would be required for each region where the method was to be applied.

Still, it might be feasible to construct smaller data bases of reference sites for localities with many mines, such as the Val D'Or region of Québec, by pooling information from individual studies at each mine site. Supplementary sampling of other reference sites would still be necessary to create a data base of useful size. If a successful model were in place, a reference areas approach would reduce the effort required for annual monitoring because control sites would no longer be needed.

A key consideration in biomonitoring with a multivariate model is how accurate are the predictions of expected community structure. Proponents of the RIVPACS model claim good success at predicting communities, with >70% of test sites correctly classified among as many as 25 different groups (Cash 1995). That level of accuracy may be satisfactory for regional surveys but it would not be sufficient for biomonitoring at individual mine sites. Moreover, models based on qualitative sampling and coarse taxonomy would not be sufficiently sensitive to detect slight or moderate impairment of benthic invertebrate communities. The only published test of the BEAST so far correctly classified 87% of sites into one of five classes defined by cluster analysis. The sediment toxicity model correctly classified 70% of the sites (Reynoldson et al. 1995).

The conclusion here is that the reference areas approach holds promise but is not yet sufficiently developed or tested for routine application at Canadian mine sites. Accurate classification of sites exposed to mine wastewaters would be necessary for the reference areas approach to be useful, and it has not yet been shown that this approach will improve detection of slight to moderate impairment. Properly chosen control sites on a river or lake already allow for most environmental variables, and including control sites from other water bodies unnecessarily adds another source of variability. However, this approach, or elements of it, may still be useful for places where ordinary controls are not possible (see Section 2.1.2), or to place the deviation of the benthic communities at affected sites into context of the full range of local variation.

2.1.2 Control Sites

The classic spatial design for assessing environmental effects of point source discharges is based on comparing the potentially affected sites against one or more control sites outside the zone of influence of the discharge. A critical question in this design is: How many control sites should there be? The simplest design has only a single control; at the other extreme would be many controls, including many controls on other water bodies, which is the essence of the reference areas approach. There is thus a continuum of choices concerning the number and location of control sites. The most efficient study design will be that which finds the best compromise between information gain and cost.

Many authors have stressed the importance of including at least two upstream control sites (Underwood 1991, 1992, Faith et al. 1995, Humphrey et al. 1995). A second upstream control, ideally located well above the first to avoid spatial correlation (Millard et al. 1985), provides a valuable measure of the natural variability between undisturbed sites along the river and thereby provides a benchmark against which downstream changes can be compared (Anderson 1990). For example, if there were a 25% difference in populations of a given mayfly species between the two control sites, this difference would provide a guideline of the magnitude of differences to expect naturally among downstream sites, irrespective of the effluent. The prudent worker would hesitate to attribute differences in abundance between control and downstream sites of <25% to the effluent, regardless of statistical significance.

A second control site acts as a back-up and a confirmation in case the first site produces unexpected results. With only a single control site the whole study could be rendered invalid if the control were found to be inappropriate, because of other, unsuspected disturbances or an unknown source of variability. In temporal designs double controls also provide information on the magnitude of change at different sites through time, and whether site differences in the absence of disturbance are approximately constant (Underwood 1991, Faith et al. 1995). For double controls to be effective, it is important that they be located a significant distance apart, both to avoid spatial correlation, and to truly reflect the range of natural variation in the water body. The distance between upstream controls should be of the same order of magnitude as the distance among downstream sites. For example, if downstream sites extend over a 5 km river reach, then the second upstream control should be located at least one or two kilometres above the first.

Double control sites are also essential for the effective application of some analytical methods. Similarity indices, for instance, require two species list for comparison, and cannot be used effectively with only one upstream site. With two sites, the similarity between control sites can be calculated to summarize the natural

variation among sites and define the magnitude of change to be considered biologically significant. Indices are then calculated comparing the nearest control site and successive downstream sites, and compared against the index for control sites. For example, if the average similarity among upstream sites was 0.80, a value of 0.70 between sites above and below the outfall could be taken as indicative of a stressed community (Pontasch and Brusven 1988).

The idea of double control sites has been taken further by some researchers who recommend several control sites (Underwood 1991, 1992, Humphrey et al. 1995). With three or more control sites, mean and variances for the control sites can be computed on the basis of sites rather than samples, providing a truer indication of the natural variation in the water body. The magnitude of effects observed at the affected sites can then be put into perspective by comparing them against the range of natural differences among control sites. This is the reasoning behind the environmental effects monitoring program for Canadian pulp mills, which recommends equal apportionment of sampling effort between sites upstream and downstream of the mill effluent outfall (DFO & Environment Canada 1995). This approach allows a balanced statistical design, based on analysis of variance using sampling stations as replicates.

Faith et al. (1995) suggest that multiple controls provide another option for confirming effects. Similarity indices could be calculated between the most affected downstream site and each control site in turn. If there really were an effluent effect, the affected site would act as an outlier and have a low similarity with all the control sites; using the two-control method mentioned above, there is always the outside chance that one of the controls is aberrant, and therefore defines a low level of "normal" inter-site similarity to compare against the downstream sites.

It has also been suggested that a control site on another nearby water body be included (Humphrey et al. 1995). The reasoning is, again, that a nearby control allows a better idea of how far the affected sites deviate from the range of variation among streams in the region (Wright et al. 1995). With the addition of more control sites on more water bodies this idea merges with the reference areas approach described earlier.

There are two considerations that militate against many control sites. First, increasing the number of control sites increases the cost of the study, and second, control sites on other rivers or lakes, if they are included in the statistical analysis, add another source of variability and probably reduce, rather than enhance, the sensitivity of the monitoring program (Smith et al. 1993). These sites may still be useful in that they allow a broader data base for comparison, but one of the assumptions of site biomonitoring is that effects within the water body receiving the wastewater are chiefly of interest.

Monitoring at mine sites may sometimes present a special case, however. Mines are often located near the headwaters of drainage basins where there are no upstream sites against which to compare sites below the mine. In that situation a control site on another stream nearby is probably the best solution, if it is understood that some sensitivity will be lost. It is critically important to match the reference stream with the mine-affected stream as closely as possible, and to find similar sites in terms of slope, discharge and substratum. Nevertheless, there is inevitably an increase in background variability with controls on a different stream and a corresponding loss of power in the study, which increases the chance that a real effect of the mine could go undetected (Humphrey et al. 1995).

Klemm et al. (1990) suggest that if a control site is established on a different watercourse, a second control site on each stream should be sampled as well. These sites would be far downstream, below the expected zone of effect on the mine-affected stream, and a corresponding distance on the reference stream. The downstream controls are a check that the two streams are comparable along the entire length under study. In the context of temporal studies, Faith et al. (1995) take this idea one step further, with the idea of a set of streams, each with an upstream and downstream site. The analysis would consist in showing that the mine-affected stream behaved as an outlier from the rest in terms of its biota at any site and in the direction of downstream trends. Unfortunately this approach would require several comparable streams, some complex statistical analysis and a great deal of work.

A more workable alternative might be to use a scaled-down version of the reference areas method. Samples from a variety of sites along several nearby streams, or at points around nearby lakes, could be used in a multivariate model to predict the expected fauna of the sites receiving mine effluent. The accuracy and sensitivity of the predictions would be limited by the number of sites and the uniformity of water bodies in the area, and there would be a large initial expense to sample many sites. If the method were successful, data from several years could be compiled into a permanent data base, and thereafter less comprehensive sampling, to demonstrate the continued validity of the model, might be sufficient. This is an idea in need of further refinement and testing.

The conclusion of this section is that two controls should be incorporated into biomonitoring studies at mine sites wherever possible. Except for special situations such as a where a tributary intercepts the river (see later), the return in information from more than two controls probably does not justify the additional

effort. Nevertheless, consideration should be given to pooling information from biomonitoring studies in regions with several mines, with a view toward building a data base for future predictive models. The problem of mine sites at headwaters is incompletely resolved and deserving of further examination.

2.1.3 Habitat Variables

Biomonitoring studies involving comparison of upstream with downstream sites constitute unreplicated natural experiments, and have been criticized on statistical grounds for "pseudoreplication" (Hurlbert 1984) in that the treatment and control sites are not selected randomly. Because the control sites must be upstream of the treatment site, it is always possible that some environmental influence other than the effluent was responsible for an observed difference in benthic invertebrate communities. For example, an influx of contaminated runoff or cold, nutrient-rich groundwater could enter the study reach between the last control site and the first downstream site, or the slope of the riverbed or canopy cover of streamside vegetation could change. Any of these extraneous influences could potentially change the structure of the benthic invertebrate community irrespective of any effluent effect. Hence, the effect of the effluent is confounded with other environmental effects, weakening statistical inference.

In simple spatial studies, two remedies to the problem of confounding have been recommended. First, care should be taken to locate sampling sites in areas that are as similar as possible with respect to key environmental variables. While it has been argued that benthic invertebrate communities downstream will naturally depart from the structure of communities upstream simply because of the progressive change in lotic ecosystems with downstream distance (Faith et al. 1991), this is probably not an issue in most streams. The structural changes predicted by the River Continuum Concept (Vannote et al. 1980) take place over the entire basin, from the extreme headwaters to the mouth of large rivers. Classification studies on river basins in the UK have demonstrated that macro-invertebrate assemblages tend to vary between rivers rather than along them (Wright et al. 1984, Ormerod 1987) because of the uniformity of water chemistry in any one river. Within any given reach of uniform order, local variations in streamside vegetation, land use, slope, depth and bottom type are far more important at determining benthos community composition than the longitudinal trend over the entire system (Corkum 1990, Resh et al. 1995). Hence, by ensuring that sampled sites are uniform in essential habitat characteristics, much of the downstream variance can be eliminated.

Naturally, even with the best efforts, there will still be some variation among sites, especially at the micro-scale of individual benthos samples. Therefore, the second tool is to measure essential habitat features at every site. These data can then be used to investigate, and possibly remove, the effects of current velocity, depth,

sediments (etc.) by regressions, blocking, analysis of covariance or similar techniques. Oddly, although most method guides recommend physical measurements at each site, and many field workers are evidently aware of the potential for standardizing data this way (DFO & Environment Canada 1995), there is no uniformity of approach. Klemm et al. (1990) and Anderson (1990) mention that physical site data should be collected, but make no mention of how the information should be used. DFO & Environment Canada (1995) discuss methods for adjusting benthos data for physical characteristics, but do not mention which variables should be measured, or how.

The physical variables of importance are those that influence the small-scale spatial distribution of invertebrates on the substratum. In discussions, experienced field biologists recommended measuring the following variables:

- water depth
- water velocity
- substratum particle size
- standing crop of algae or detritus (running water) or total organic carbon (standing water).

Elevation and slope may be important in mountain streams (Corkum and Ciborowski 1988). Some or all of these variables are normally measured as a standard part of biomonitoring surveys. A distinction must be made here between routine water quality measurements and physical data collected specifically for the purpose of quantifying habitats. Water quality variables such as temperature, pH, dissolved oxygen, conductivity, nutrients and turbidity are useful in understanding the setting and the nature of the disturbance, but they are less likely to correlate with the microdistribution of benthic invertebrates. Similarly, variables such as canopy cover, bank stability, and nearby land use may be useful but are more likely to be correlated with invertebrate densities on a basin-wide scale (Pettigrove 1990).

Measurements of depth and water velocity are straightforward and require only a staff or sounding line and a velocity meter. Particle size of the bottom material can be measured by sieving samples, but rough visual estimates will work as well (Fernet and Walder 1986). Determining the particle size distribution of benthic sediments is not often difficult in streams and shallow rivers, but it can be challenging in lakes or turbid waters, especially where bottom materials are very heterogenous. Estimates of organic detritus are easily obtained by weighing the oven-dried residue from sorted invertebrate samples before and after burning in a muffle furnace to correct for inorganic sediments. Rock scrapes can be used to measure benthic algae, but here again many workers rely on qualitative classifications based on field inspection. Methods for these

procedures are available in any methods manual such as APHA (1992). The recent text on methods in stream ecology by Hauer and Lamberti (1996) provides solid basic instruction on physical measurements.

The strength of correlations between physical variables and species densities will be much improved if data can be obtained for individual samples instead of the site as a whole (assuming here that one is using the approach of several replicates at each site). This appears to be especially important for current velocity, of which minor variations from one sample to the next can strongly influence catches of current-sensitive mayflies (personal observation). Physical data for every sample might not be feasible if the number of samples per site were larger (see Section 3.1) but in that case a few measurements for the entire sampling site would suffice for data adjustment.

While it would seem logical that data adjustments for physical habitat features would be a powerful technique for reducing sampling variability, in practice these methods must be applied with caution, especially if there is a downstream trend that coincides with the expected disturbance gradient, leading to confounding of the two. In that situation no correction for the measured habitat variable is possible. Where a variable such as depth or current velocity varies substantially among sites, the effect of adjustments for these covariates on sensitive species can be dramatic and data adjustments should be applied circumspectly. Some biologists recommend analysing the data twice, with and without the habitat adjustment. There is a paucity of published work on the benefits of physical habitat adjustments, and better guidance on the best way to go about it would be welcome.

2.1.4 Multiple Contaminant Sources

The discussion to this point has assumed the simplest situation, where a single point-source effluent or similar local disturbance is affecting a uniform reach of stream. At many mine sites, perhaps most, there are several to many sources of potential contamination, and both point sources and nonpoint sources may be in evidence. Moreover, there may be other confounding factors that influence the health of the receiving water body independent of the mine in question. Sites upstream from the mine may already be impaired by other pollution sources, including other mines. Downstream sites within the zone of potential effect of mine effluent or run-off may be insulted by pollutants from other mine sites, nearby industries, or variously treated domestic sewage from nearby communities or even the mine housing itself. Finally, tributaries carrying water of similar or different background quality, but without the contaminants contributed by mine wastewaters, may enter the receiving stream at any point.

The problem of multiple contaminant sources is widespread in benthic invertebrate monitoring, but there has been little research on optimization of sampling to deal with it. While benthos sampling downstream of the last outfall will measure the combined effect of all the effluents, it is presently not possible to assign the effect proportionately to one source or another. Invertebrate populations at any site will reflect the sum of all stresses bearing upon the community at that site, regardless of their source, and the degree of effect attributable to one or another effluent is difficult to unravel, especially given complex effluents and the possibility of synergisms.

A number of solutions to this problem have been proposed. The most direct solution is to sample above and below each effluent or contaminant source in the series. The difference in benthic community composition between successive sites indicates the effect of any contaminants entering between them; in effect, each site acts as a control for the one below it. Such an approach was used with some success to separate effects of numerous effluents on the North Saskatchewan River (Golder Associates 1993). Tributaries present a comparable challenge. Bournaud et al. (1996), established sample sites below every tributary on the Rhône River to gauge their effect on downstream trends in the main stem.

In practice, this approach is only partially successful. The difficulty is that benthic communities tend to reflect the effect of the strongest effluent. Where one effluent upstream is strongly detrimental to invertebrate populations, smaller effects downstream are difficult to see because the sensitive species that would illustrate the effect are already lost. In the Saskatchewan River example, the effect of enrichment from Edmonton's sewage effluent was so pervasive that the small, local effects of industrial effluents could barely be detected. Similarly, while a succession of small pressures on the community should be visible as a progressive change in community structure downstream, in practice the changes below the first effluent are difficult to detect.

Another possibility is to apportion the effects according to the strength of the various effluents. In this approach, the largest effluent, in terms of chemical concentration and volume after dilution, would be assumed to be responsible for the greatest proportion of the effect, while smaller effects would be ascribed to less potent or voluminous effluents. This idea encounters both theoretical and practical difficulties. Quantitative comparisons among effluents are feasible where the effluents are all of similar composition (e.g., wastewaters from several metals mines) but become problematic where effluents of widely divergent character are involved (e.g., a metal mine effluent and sewage effluent). It could be argued as well that the apportionment of effects is based on supposition rather than observations, and as outlined above, the responses of benthic populations in the field may be more complex.

Better resolution might be possible by combining invertebrate sampling with toxicity tests on the individual effluents, as was done for the Saskatchewan River Study. Sediment toxicity tests, effluent plume delineations and tracer studies can also be used to sort out the various influences on the benthos. Not every routine study has the budget for such complex analyses. Alternatively, benthos at the affected sites can be compared against pristine sites in a nearby drainage, as discussed earlier, if other conditions can reasonably be assumed to be comparable.

To determine the presence, nature and extent of all the impairment and recovery zones in a river receiving multiple effluents with present techniques remains a challenge. At mine sites, the problem is further exacerbated because nonpoint-source contaminants from overburden, disturbed ground and tailings are likely to be coincident with, and as important as, point-source effluents. Multiple contaminant sources are a difficult problem in need of further research.

2.2 Sequential Decision Plans

2.2.1 Overview

Sequential decision plans, also known as sequential sampling plans or sequential analysis plans, are a method of biomonitoring that can drastically reduce the number of samples required to detect impairment in a biomonitoring program. In a sequential decision plan, the number of samples to be analyzed is not fixed in advance; rather, samples are collected, sorted and analyzed sequentially until a decision can be made to classify a site as impaired or unimpaired according to predetermined levels of risk and precision. The attraction of these plans is that they can reduce the cost of biomonitoring by 50-60%, while still allowing clear-cut decisions as to whether degradation is or is not occurring (Resh and Price 1984).

Sequential decision plans have been used for many years in manufacturing, and in a variety of scientific fields, including marine biology, environmental microbiology, and especially terrestrial pest control, but have seldom been applied to benthic invertebrate monitoring. Dr. V.H. Resh has championed the application of sequential decision plans to benthos biomonitoring in a series of publications (Resh and Price 1984, Resh et al. 1988, Jackson and Resh 1988, 1989), on which the discussion here is founded. The general utility of the sequential decision approach for monitoring effects of mines or other industries on fresh waters remains untested.

The central assumption of a sequential decision plan is that means and variances of species populations are not of interest in themselves; what matters is the ability to detect the effects of pollution on those populations,

which is equivalent to classifying sites as impaired or unimpaired. (The term *impaired* is used here to describe benthic communities disturbed by pollution or other human actions, in preference to *impacted*, an undeservedly popular atrocity.) No statistical parameters are calculated in a sequential comparison plan, and no hypothesis-based tests (ANOVA, *t*-test) are carried out. Instead, the data from each sample are combined with data from all previous samples to classify the site into one of three categories: impaired, not impaired or no-decision. Fewer samples are examined with a sequential decision plan than with ordinary parametric inference because sample analysis stops when the minimum information needed to classify the site as impaired or unimpaired is obtained (Jackson and Resh 1988).

The essence of a sequential decision plan is a graph or table (Figure 1A and Table 1) consisting of two decision lines representing cumulative totals for a study variable such as mean species richness or population density of a species of interest. The decision lines define the three regions of impaired, unimpaired and no-decision or continue. The site is classified by examining each sample in sequence, and adding the observed count or measure to the cumulative total for all previous samples. The result is then plotted on the graph (or compared with the equivalent table) and compared against the decision lines. If the point lies above the upper decision line the site, or more properly the benthic community, is declared impaired and no more samples are examined; if the point lies below the lower decision line the site is declared unimpaired and again sample sorting ends; if the point lies between the two lines, in the no-decision zone, another sample is examined and the process is repeated. The sequence continues until the site is classified or all samples collected have been examined (Jackson and Resh 1989).

This process is illustrated in Table 1 (from Jackson and Resh 1989), where the decision lines might represent population densities for a common benthic taxon. If the first sample contained 90 individuals, the point falls in the no-decision zone and sampling continues. A second sample with 40 individuals would lead to a cumulative total of 130 and still no decision. A third sample containing 50 individuals would bring the cumulative total to 180, sufficient to classify the site as unimpaired. No further samples would be examined.

2.2.2 Designing A Plan

The decision lines in a sequential decision plan express hypotheses about the nature of the biological response to disturbance. These hypotheses may be derived from comparison of clean and polluted sites, from laboratory or field studies on the effect of a chemical or effluent, or from background knowledge of the ecology of the species of interest and their populations at reference sites. Decision lines can be based on either an increase or a decrease in population density, or any other variable of interest.

In the simplest case, impairment might be defined as a set deviation from the mean population at the control site. For example, if mean density of caddisflies in the genus *Hydropsyche* at control sites was 200 individuals per sample, and an effluent was expected to create inhospitable conditions for these insects, then a density substantially less than 200 per sample would be indicative of impairment. However, given sampling variability, a sample could contain fewer than 200 individuals and still represent an unimpaired site. This uncertainty would be incorporated by adjusting the decision line for the standard error of the mean. Resh et al. (1988) suggest that (Mean - 2 SE) is a reasonable, conservative value, i.e., any sample that is no more than two standard errors less than the mean should not be considered indicative of impairment. If the standard error were 25 in the above example, the impairment decision line would be defined by these points: 1 sample, 150; 2 samples, 300; 3 samples 450; etc.

Where one or more impaired sites are available for comparison, the no-impairment decision line could be set in the same way, as (Mean + 2 SE). In the absence of measured impaired sites, probably the more typical case, the lower decision line can be defined by deciding what magnitude of change (e.g., 50%, 75%) in the population can be detected with a reasonable number of samples. Uncertainty would again be included by adding the standard error or a multiple of it to the predicted impaired-site mean. Occasionally, experimental data might allow more precise decision lines. Resh and Price (1984) describe a plan in which the expected changes in population density of the chironomid *Cricotopus* spp. in response to petroleum were determined from population counts on experimentally oiled artificial substrates. Another example (Resh et al. 1988) incorporates annual variation in population density of hydropsychid caddisflies according to rainfall.

Sequential decision plans explicitly consider the risk of error, which in a plan consists in a probability of misclassification. A Type I error, of which the probability equals α, occurs when a site is classified as impaired when in fact it is unimpaired; in common parlance a Type I error is a False Positive. Conversely, a Type II error or False Negative occurs when an impaired site is incorrectly classified as an unimpaired site. The probability of a Type II error is represented by ß.

In a sequential decision plan, acceptable levels of both Type I and Type II errors are decided beforehand. The acceptable error rates are set based on the consequences of making a mistake and the practicality of collecting the number of samples required. A Type I error, incorrectly declaring a site impaired, could have economic implications if expensive remedial works or corrective actions were taken that were not necessary. Failing to recognize impairment (Type II error) on the other hand, can be costly to the environment because the degradation would not be recognized until the next monitoring period, and more severe degradation could occur in the meantime. Jackson and Resh (1988, 1989) recommend setting α and ß conservatively at 0.05, but other

values should be considered depending on the circumstances. The key point is that the risk of both Type I and Type II errors is always known in a sequential decision plan, and modifications to the plan take into account effects on both accuracy and power.

Once the levels of α and ß have been decided, the next step is to determine the best mathematical model to describe the distribution of the variable of interest. Sequential decision plans can be developed for data conforming to the normal, binomial, negative binomial or Poisson distributions (Jackson and Resh 1988). As discussed in Section 3.1, population data for individual species or total numbers of individuals in a sample are often approximated by the negative binomial distribution; species richness, on the other hand, conforms more closely to a Poisson distribution (Resh et al. 1988), and some measures of species diversity follow the normal distribution (Jackson and Resh 1989).

The decision lines at the heart of a sequential decision plan are linear equations relating cumulative totals of the count or measure of interest against number of replicate samples. The slopes and intercepts of the line are calculated using the predecided classification thresholds (the limits defining what results are indicative of an impaired or unimpaired site), the error limits α and ß, and, where the normal or negative binomial distribution applies, the sample variance or the dispersion constant k. With this information, two other features of the decision plan can be calculated: the operating characteristic curve and the average sample number curve.

The operating characteristic curve (Figure 1B) plots the probability of classifying the site as impaired or unimpaired, equivalent to accepting a hypothesis defined by the decision lines, against the mean of the variable of interest. Of course, only one of the competing hypotheses is correct. The values of α and ß delimit the tails of the probability distribution within which a classification of the site one way or the other would be within the acceptable limits of error. The area between the two tails defines the no-decision region in the plan.

The operating characteristic curve in turn is used to define the average sample number curve, which plots the number of samples that will have to be examined, on average, to make a decision under the plan, against the mean count of the variable of interest in each sample (Figure 1C). These plots tend to increase with increasing counts of the variable, and then decline again, because either very low counts or very high counts will lead to the site being quickly classified one way or the other. Intermediate counts lead to ambiguous results (no decision) so a larger number of samples is necessary to classify the site. The average sample number curve can be used to decide the maximum number of samples to collect in the field.

If sequential decision plans were included in biomonitoring at active mines, the logical procedure would be to use information from control sites to determine the unimpaired condition. Jackson and Resh (1989) suggest that all the samples from the control site would be sorted as usual, and the information on population density would be used to set the decision lines. Only the samples necessary for a decision would be sorted at downstream sites. There is also the attractive possibility of using sequential decision plans to test hypotheses from other facets of the monitoring program, such as toxicity tests on wastewaters or heavy metals, perhaps combined with the known or expected distribution of metal contamination at study sites.

2.2.3 Advantages and Disadvantages

There is no question that, when the objective is to detect a defined impairment at a given level of precision and accuracy, sequential comparison plans can produce huge time savings compared with the conventional approach. Using data for a California stream intensely sampled with the Surber sampler, Jackson and Resh (1989) estimated the number of samples needed to classify the undisturbed site correctly, based on species richness, species diversity (inverse of Simpson's Dominance index) and population density of the abundant mayfly *Cinygmula*. The comparison was repeated for decision lines based on reductions in the test variable of 10% through 60%. For species richness (Poisson distribution), effort, defined as time taken to sort samples, was reduced by 50-60% compared with conventional analysis. For diversity (normal distribution), the reduction was 60-78%, and for *Cinygmula* (negative binomial distribution) the savings was near 50% in every test.

Because it was known in this test that the sampled site was pristine, the proportion of misclassifications (labelling the site impaired) or failures to classify (remaining in the no-decision zone when the maximum of 15 samples had been sorted) could also be calculated. The site was incorrectly classified as impaired in 5% of the simulations or less, regardless of the decision line used, with species diversity or *Cinygmula* population density as variables. The simulations with species richness never produced a misclassification. However, with decision lines representing 20% through 60% reductions in the variables, from 5% to 25% of the simulations resulted in no classification when all 15 samples were included. The decision lines based on 10% reductions lead to no classification 50-65% of the time, suggesting that a 10% change is below the detection threshold of the method.

In situations were a definite classification decision is necessary, a technique known as the truncated decision method is available. This step consists of nothing more than bisecting the no-decision zone, and classifying sites according to the half of the region in which they lie. Applying truncation to the California stream data

lead to correct classifications in 90% to 100% of the simulations for all variables. Hence, in this example a clean stream had as much as a 10% chance of being labelled impaired, rather more than would be expected from $\alpha = 0.05$. For routine biomonitoring, it would probably be better not to apply truncation and rely on other methods, or more sampling, to decide if unclassified sites suffer impairment.

A number of important assumptions must be met before sequential decision plans can be used. First, the sampling distribution of the variable of interest must be known. Equations for decision lines are available only for the four types of distribution mentioned earlier. Further, the parameters of the distribution, such as the variance or the dispersion coefficient, must be known.

Second, sequential decision plans require a one-way hypothesis about the effect of the disturbance on the study variable. They cannot be used to explore the effect of an effluent or other perturbation on the benthic community, but only to determine objectively whether a particular effect has occurred. Though a number of data sources are available on which to found hypotheses, the inability of these plans to detect unanticipated change is a real limitation.

The third assumption is that the time, and therefore cost, of collecting samples in the field is only a small fraction of the time taken to sort samples in the laboratory and identify the specimens in them. As discussed in Section 3.1, this assumption is almost invariably true, especially when samplers of conventional size are used (Resh and Price 1984). If sampling schemes were to move from large to small samplers, with an increased number of replicates, the validity of this assumption would need to be re-evaluated.

In addition to the need to meet some restrictive assumptions, sequential decision plans have other limitations:

(1) They use only one variable to make a decision about site classification. If a community-level variable such as a diversity index or similarity measure was used, then information about all species would in some way be included. But plans based on one taxon or one variable, like species richness, ignore all the other information contained in the sample. On the other hand, to make a weight-of-evidence argument based on results for many species or variables would be tedious and complicated, because all the steps of the decision plan would need to be repeated for each variable.

(2) They do not always lead to a definite answer. As the earlier example illustrates, when the total number of samples is limited there is no guarantee that a sequential decision plan will be able to classify every site. It is a waste of effort to sample and sort any number of benthos samples if the information they provide about

environmental quality is ambiguous. In practice the number of unclassified sites will be few and can be reduced by changing α or ß.

(3) They do not make an entirely objective analysis. To an extent this is true of any statistical test, because the significance level and the number of samples, which influences power, must be chosen by the investigator. However, in sequential decision plans a decision must also be made regarding the magnitude of a difference to be considered indicative of impairment. Questions about what constitutes a "significant" effect in biological surveillance, and who should make the determination, are always controversial.

Variance in field populations is not an inherent part of sequential decision plans (Jackson and Resh 1988). The criteria of (Mean ± 2 SE) as a basis for decision lines based on reference areas is simply a choice favoured by one investigator (Resh et al. 1988, Jackson and Resh 1989), and has no theoretical basis. Therefore, a limit based on (Mean ± 1 SE) or (Mean ± 3 SE) or some other choice such as a 50% or 100% reduction in population density, would be equally valid, although they would lead to widely different site classifications.

(4) They say nothing about the severity of the effect. Sequential decision plans classify sites as impaired or unimpaired; they do not describe the nature or severity of the impairment. It follows from this that they will not detect an effect that is different than the one for which they were designed. For example, a plan based on metal toxicity will not detect effects of organic enrichment. However, a refinement of the basic plan that will detect gradients of impairment is possible (Jackson and Resh 1989). The impaired designation can be subdivided into smaller classes representing moderate or severe impairment. The decision plan then has four decision lines, and requires two sequential decisions. The first separates severely impaired sites from sites of moderate or no impairment. The remaining sites are then separated, with information from more samples, into moderately impaired and unimpaired sites. A graphical example of such a plan is shown in Figure 2.

If the potential of decision plans is to be realized, a change in the way sample processing is currently handled would be required. Many consultants subcontract sample sorting and specimen identifications to specialists who charge a fixed price per sample or for the lot. When all the samples are finished, the investigator receives a batch of species lists that form the basis for subsequent analysis. Sequential decision plans would require much closer collaboration between analysts and taxonomists, and might also require modifications of the logistics and pricing structure of taxonomic work. The effect of other possible procedural changes, particularly reduction in sample size, on decision plans is unknown.

Sequential decision plans are a potentially useful idea that could lead to substantial cost savings for biomonitoring under the right circumstances. The limitations described above must be overcome; nevertheless the potential improvement in efficiency offered by these plans is so great that further examination of their place in biomonitoring at mine sites would seem appropriate.

2.3 Rapid Assessment Approaches

Rapid assessment approaches to biomonitoring are mentioned here for completeness because there has been a great deal of interest and research in these methods in the past ten years (see Resh and Jackson 1993 and Resh et al. 1995 for reviews). Rapid assessment procedures are now widely used by state and federal agencies in the United States (Plafkin et al. 1989, Barbour et al. 1992, 1996) and the United Kingdom (Wright et al. 1988), and are now being developed in Australia (Chessman 1995, Growns et al. 1995). However, despite their utility for regional monitoring, it is unlikely that these quick-evaluation methods would be of great value for biomonitoring at Canadian mine sites.

Rapid assessment approaches are designed to identify water quality problems associated with point-source and nonpoint-source pollution or other anthropogenic perturbations and to document long-term changes in water quality within a region. Hence, they are based on comparisons between surveyed sites and clean reference sites that are taken as representative of the natural condition in the absence of human influence. A second objective of these methods is to summarize results of site surveys in a way that can be easily understood by non-specialists such as managers, politicians and the concerned public. This objective is accomplished by summarizing conditions at a site as a single-number score that expresses the health of the system on a relative scale (as "good", "slightly degraded", "poor" etc.) in comparison with the regional reference sites. Rapid assessment procedures have been promoted by regulatory agencies like the United States Environmental Protection Agency (USEPA) and similar state or regional agencies that needed a method to assess water quality in thousands of kilometres of flowing waters extending across a vast geographic area. These procedures are designed to quickly screen large regions, pinpointing trouble spots for more detailed investigation.

The biomonitoring method in which potentially impaired sites are compared against regional reference sites or predicted benthic community structure derived from those sites is described in Section 2.1.1. The reference areas method is integral to the application of rapid assessment approaches. Notwithstanding, the reference areas approach can equally be applied to any kind of quantitative or qualitative sampling, and does not depend on rapid assessment procedures. The evaluation of rapid assessment approaches in this section applies only to the sampling methods themselves, not to the broader reference areas approach in which they are imbedded.

Rapid assessment procedures sharply reduce the cost associated with a biomonitoring program compared with traditional quantitative approaches by employing some or all of these time-saving measures: (1) the number of replicate samples taken and the variety of habitats sampled is reduced; (2) only a fraction of the animals in the sample are considered, which speeds sorting and identification; (3) identifications to genus, family or even higher levels are used; and (4) standard, simple measures of community composition, termed *metrics*, are used in place of statistical comparisons (Resh and Jackson 1993). Most methods are applicable only to wadable streams and rivers.

In a typical rapid assessment protocol, about which there are innumerable variations, a single pooled sample would be collected at a site with a D-net or kicknet in a set time, say 20 minutes. Sampling would either concentrate exclusively on riffles or effort would be apportioned over all habitats, riffles, pools, organic debris and stream margins. Either the entire sample or some fixed proportion of it (first 100, 200 or 300 animals) would be sorted and identified. The appropriate metrics would then be calculated and the results compared against the standard scale or data for regional reference sites. In the US, rapid assessment approaches have been designed to go from field sampling to final report in as few as five working days (Resh et al. 1995).

As discussed earlier, (Section 2.1.1) two approaches to biomonitoring have independently developed in the United States and UK. The American approach incorporates regional reference sites as the basis for comparison with study streams. For each region within a state or other jurisdiction, particular streams or stream reaches are selected that are thought to best exemplify the environmental conditions that would obtain throughout the regions in the absence of human influence. These streams must be undisturbed by municipal or industrial effluents, agricultural runoff, forestry or land clearing, and must be physically and chemically comparable with the other streams in the region. The benthic communities at the reference sites are then used as the standard defining best water quality, and other sites are ranked according to how closely their community composition matches that at the reference sites. In the UK, extensive sampling of every river system on the island was undertaken, and the results used in a multivariate analysis to ordinate the sites and determine the key environmental variables controlling benthic community composition. This model is then used to predict the fauna that should occur at a test site if it were free from human disturbance (Wright et al. 1984, Armitage et al. 1987, Moss et al. 1987).

Likewise, the procedures used to rank sites in the American system are of two types. Some systems compute a single index of water quality, either borrowing or adapting well-known biotic indices or developing new ones (e.g., Chessman 1995). The Biotic Index developed by Hilsenhoff (1987, 1988) for Wisconsin, which ranks sites according to the mean water quality tolerances of invertebrate species or families, is perhaps the best

known example of this approach. More often, a *multimetric* approach is used, in which many separate metrics, including biotic indices among them, are separately evaluated. The redundancy built into a multimetric approach reduces the risk of misclassifying a site based on random error in one measurement. The results of all the metrics are often combined into a single index that expresses the overall condition of the site (Barbour et al. 1996).

The number and variety of metrics that have been proposed borders on bewildering. Some of the more common, and more successful, are (Resh and Jackson 1993, Barbour et al. 1996):

(1) number of taxa,
(2) number of individuals,
(3) number of taxa in the Ephemeroptera, Plecoptera and Trichoptera (EPT richness),
(4) similarity indices (per cent similarity, Jaccard Coefficient, Margalef's Index, and many others)
(5) Biotic Indices,
(6) per cent dominant taxa,
(7) ratios of community composition (hydropsychids to total Trichoptera, EPT taxa to chironomids, Tanytarsini as a percentage of total Chironomidae, etc.), and
(8) functional feeding groups (percentages of scrapers, predators, shredders, filterers).

This list is representative only. Resh and Jackson (1993) provide a somewhat more comprehensive list that runs to eight pages. Many of these metrics are the same as those that would be used in conventional parametric comparison of sites. The difference is in how the data are collected and analyzed.

Many rapid assessment procedures incorporate a hierarchial structure of detail. A different level of survey intensity can be chosen according to the objectives of the study. For example, the USEPA method establishes three levels of benthic invertebrate sampling: level I is a reconnaissance survey to document the presence of obvious impairment and to see if more detailed studies are necessary; levels II and III are for site rankings, with different levels of effort and expertise in each (Plafkin et al. 1989).

The American system is also unique in that it attempts to considers habitat degradation as well as water quality deterioration as a cause of benthic community impairment. A short list of variables reflecting suitability of habitat conditions are assigned numerical scores based on visual inspection or a minimal amount of measurement. The list includes factors such as bank stability and erosion potential, riparian vegetation type and cover, channel morphology (ratios among pools and riffles, runs and bends), and microhabitat features such as sediment particle size and stability. The data from the habitat assessment are used in the assessment stage

to decide whether poor or instable habitat may be contributing to the benthic invertebrate community structure (Resh et al. 1995).

The price of rapidity and accessibility in rapid assessment approaches is a loss of accuracy. Rapid assessment approaches have been likened to a thermometer, used to "take the temperature" of an aquatic ecosystem. A deviation from the expected setpoint, if we can define it, indicates that ecosystem health is impaired, and that further investigation is needed (Resh and Jackson 1993). Because of the lack of replication, absence of statistical comparisons and reliance on simple counts, the sensitivity of rapid assessment approaches is severely limited. Metrics based on taxa richness for the whole sample or within a particular group are biased and inaccurate when only a fixed number of animals is sorted from the sample (Courtemanch 1996). Tests of these methods at individual sites have shown both failure to detect moderate levels of known disturbance and incorrect warnings of impairment at pristine sites (Resh and Jackson 1993). Hence, while these methods are suitable for regional comparisons and quick evaluation of severe degradation, they are neither sensitive nor robust enough to replace replicated statistical approaches (Kerans et al. 1992).

Nevertheless, perhaps some of the concepts underpinning rapid assessment approaches can be applied to regular biomonitoring at mine sites. The multimetric approach is certainly adaptable to statistical comparisons, especially with multivariate techniques, and research on metrics for rapid assessment approaches can reveal which metrics are sensitive and robust and which are too noisy or redundant (Barbour et al. 1992, 1996). Rapid assessment techniques may be suitable for background monitoring at reference sites to estimate annual variation in benthic communities (Armitage and Gunn 1996). Finally, it might be possible to incorporate a hierarchial approach like that in the USEPA rapid assessment method (Plafkin et al. 1989) into some kinds of routine monitoring. For example, in a river reach suffering serious degradation, simple surveys to confirm that condition might suffice until remedial works are finished or the source of contamination is removed. When the simple survey could no longer detect impairment, a statistically based, replicated study would be done to more carefully evaluate whether the site had recovered. Cost efficiency is not necessarily served by using sensitive tools where the ecosystem impairment is obvious.

3. Field Methods

3.1 Sample Size and Replication

3.1.1 Overview

The effect of sampler size on sampling efficiency and cost has been examined by a number of researchers in both marine and freshwater environments. The uniform conclusion is that where sampling cost is small compared with processing costs (sorting and enumeration) there is a clear advantage to decreasing sampler size and increasing replication (Downing 1979, Resh 1979, Pringle 1984, Morin 1985, Ferraro et al. 1989, 1994). In benthic invertebrate studies, the cost of sample collection is usually a small fraction of processing costs. Resh and Price (1984) reported from a survey of consultants and researchers that sorting and identification time in the laboratory constituted well over 90% of the total time in a benthic survey. Under such circumstances, the advantages of numerous small samples are clear. Most standard sampling devices, such as the Surber sampler, or Hess or Neill cylinder samplers, take samples that are far too large. Smaller samples can be sorted more quickly, and the saved effort can then be expended on collection of more replicates, which improves the precision of population density estimates.

3.1.2 Spatial Distribution of Benthic Invertebrates

In the present discussion, *sample size* refers to the area of substratum sampled by the sampling device; the number of samples collected at a site (*n* in statistical parlance) is *replication.* The general rules for selecting the number of samples necessary to achieve a certain precision (i.e., standard error of the mean) in a benthos sampling program are generally well known, though they are not universally applied. The standard treatments by Elliott (1977) and Green (1979) are widely quoted. These calculations require information on the expected mean and variance of the population being sampled, usually determined from experience or preliminary studies.

The fundamental difficulty in sampling benthic invertebrates arises from the non-random distribution of organisms on the river bottom. This aggregated distribution evidently arises from animals actively selecting microsites that are favourable in terms of current velocity, food resources or safety from predators (Resh 1979). For example, net-spinning caddisflies in the family Hydropsychidae orient themselves toward the current to optimize food capture, and different species select different locations according to current velocity near the rock surface (Williams and Hynes 1973). However, even where the substratum is apparently uniform,

aggregated distributions are still found (Shipley 1987). Aggregation has been demonstrated in the benthic fauna of lakes and large rivers (Downing 1979, Veijola et al. 1996) in streams (Resh 1979, Morin 1985) and in the ocean (Vézina 1988) and appears to be a universal feature among benthic invertebrate populations.

Resh (1979) demonstrated the effect of spatial variability among stream insects by comparing densities of a common caddisfly, *Cheumatopsyche pettiti* in 26 pairs of Surber samples taken side by side in a uniform stream riffle. If the distribution of organisms were more or less uniform over a large scale (relative to the sampler), then the number of caddisflies in adjacent samples would be about the same. In fact, equal numbers of individuals in adjacent samples were rarely found; while similar numbers did occur in some samples, in others the counts were quite different (Figure 3). Similarly aggregated populations are found in all groups of benthic invertebrates (slightly less in predators, slightly more in filter-feeders; Morin 1985) and in the community as a whole (Downing 1979).

3.1.3 Implications of Aggregation for Sample Size

The net effect of aggregated species distributions is that variance is greater than would be expected based on a normal distribution and consequently a large number of samples is necessary to estimate population means with a reasonable degree of precision. Estimation of sampling requirements is complicated when the distribution departs from normality because the degree of aggregation of the population strongly affects the sampling intensity necessary for a given precision. There has been debate over which distribution is the best model to describe benthic invertebrates, with the negative binomial being the most often cited (Resh and Price 1984, Resh et al. 1988).

Downing (1979) used data from 23 studies of lakes and large rivers to derive an empirical function relating the sample mean from replicate benthos samples to the sample variance, and thus side-stepped the issue of the best theoretical model. His regression showed that the standard deviation of a set of replicates, with populations expressed as numbers per square metre, was predictable from the sample mean (the variance increases as the mean increases) and the capture area of the sampling device. By substituting desired levels of precision, such as a standard error of 20%, into the regression, it is possible to predict the most probable number of samples of any size that must be collected to achieve the desired precision, for any given density of animals.

Exemplary results of Downing's equation, for a standard error of 20% of the mean, are reproduced here as Table 2. (The table has been corrected for miscalculations as reported in Downing (1980) and Riddle (1989).)

Two clear trends emerge. To achieve the same precision, (1) more samples must be taken with small samplers than with large samplers, and (2) more samples must be taken at low density than at high density.

However, to evaluate efficiency, the cost of sorting and enumerating samples must be considered. The time taken to sort a sample varies with the amount of detritus and the number of invertebrates in it, which are determined by the area of the sampler, so sampler area can provide a comparative value for the cost of sample processing. Multiplying the number of replicates in Table 2 by the area of the sampler generates an estimate of the relative cost-efficiencies of different combinations of replication and sampler size. For ease of comparison, these estimates are expressed in Table 3 as a proportion of the area of sediment that must be sorted in a sample of 1000 cm^2, roughly the area of a Surber sampler or Neill cylinder sampler.

The cost benefits of taking a larger number of smaller samples are now apparent. Depending on the population density, the cost of processing samples of 100 cm^2 would be one third to one tenth the cost of processing 1000-cm^2 samples, with the same precision. At high population densities, where a smaller sampler could be used, the benefit is even greater, up to fifty times for a 20-cm^2 sampler, though in reality only soft-sediment corers could be that small. Hence, Downing (1979, 1989) concludes that small samples are generally much more cost efficient than large samplers for benthic invertebrate sampling. Resh (1979) remarks that the prevalence of large-area samplers and minimal replication (usually 5 or less) in the freshwater biology literature accounts for the high variability typically reported for population estimates of stream benthos.

Although Downing's (1979) work was based on lakes and large rivers, later studies have shown that the same principles apply to the benthos of streams (Morin 1985) and coastal marine environments (Shipley 1987, Vézina 1988) and for that matter to sampling of epiphytic organisms (Downing and Cyr 1985), and even aquatic plants (Downing and Anderson 1985) and seaweeds (Pringle 1984). Morin's (1985) review based on data from 19 studies of stream benthos found that aggregation was even stronger among stream-dwelling organisms than in lakes, but the same relationships among population density, sampler size and precision reported by Downing (1979) emerged.

Morin's key results are presented in Figure 4. For a given sampler, the number of replicates needed increases with the desired precision and decreases with increasing mean density of invertebrates (Figure 4A). When the precision is specified, such as a standard error of 20%, the number of replicates needed decreases with increasing size of the sampler (Figure 4B), but the total surface area sampled, and therefore the amount of detritus that must be sorted, increases (Figure 4C). This analysis again leads to the conclusion that where

processing time is a large part of sample collection cost relative to field time, the most cost-efficient scheme is to use the smallest sampler possible.

The improvement in cost-efficiency obtained from a reduction in sample size can be used both to decrease the total sampling cost, or to increase the precision of population density estimates, especially for less abundant species that would normally be undersampled. As an example, Mackie and Bailey (1981) describe a simple stream-bottom sampler, called a T-sampler, that samples an area of 100 cm^2, as compared with 930 cm^2 for a standard square-foot Surber sampler. In tests in a productive river, the T-sampler collected significantly greater numbers of total organisms and numbers of the numerically dominant taxa (on an areal basis), and was equally efficient for other species. But approximately 6-7 T-samples could be collected and sorted in the time taken to process just one Surber sample. Hence, large numbers of replicates (>30) could be taken at a site with the same effort presently expended for five replicates using square-foot cylinder samples. The increase in replication far outweighs the smaller number of individuals in each sample.

The discussion thus far has assumed a conventional sampling scheme in which individual sites are the basis for spatial comparisons and replicate samples are collected at each site. In the study design used in the pulp mill environmental effects monitoring program, comparisons are made among larger areas, with individual sites within them serving as replicates. In this design a single site may be represented by only one sample, and the argument about many small samples becomes a question of how many subsamples to include in the pooled site sample.

Even in this design there remain compelling reasons for using a larger number of small samples in place of a single large one. In addition to the statistical justification presented earlier, small samples provide more complete coverage of the site. Small samples integrate effects of fine-scale variation in habitat characteristics that so strongly influence the distribution of benthic animals, and therefore are more likely to return a truly representative sample of the site fauna than a single large sample at one point. Where the substratum is heterogenous, for example, a single sample may by random chance be taken from a point of very low or very high animal density (for the whole sample or any taxon of interest), whereas several small samples are much more likely to include the full range of densities. A parallel argument applies to other habitat characteristics that vary according to location in the channel, such as current velocity, detritus accumulations, canopy cover or algal biomass. Hence even where the sampling effort at an individual site is limited, numerous small samples are a superior choice to a single large sample.

3.1.4 Minimum Sample Size

The improvement in efficiency of smaller samples is greatest when population density in the sampled stream is greatest. For sparse populations, such as those in Rocky Mountain streams or northern bog drainages, large sample sizes are competitive in cost efficiency with smaller sizes (Morin 1985, Riddle 1989). Moreover, because benthic animals are discrete units, there is a lower limit to effective sampler size for any population, at the point where most samples contain no individuals and a few samples contain one or two. For combinations of very small sample size and very low population density (mean density <0.5 per sample), a standard error of 20% or less cannot be achieved with any practical number of samples (Riddle 1989, Downing 1989). Given the distribution of species in benthic communities (a few common species and many uncommon or rare species) there will always be some rare species that will be below the effective size threshold of the sampler. However, because smaller total area sampled by a small sampler can be more than compensated by increased replication, the number of uncommon species in the list for which a density estimate of given precision can be obtained will be at least as great as with a large-area sampler.

The implicit assumption underlying this analysis is that the time required to collect samples in the field is a small fraction of the time required to sort and enumerate them, so reduction in laboratory time is the critical factor in reducing sampling cost. While firm estimates of sorting time are elusive, in most studies time in the field constitutes only about 5% of the sample processing time (Resh and Price 1984). Given the overwhelming dominance of laboratory time in sampling costs, the potential for improvement in cost-efficiency from reducing sampler area and increasing replication is considerable.

However, while conventional samples require far more laboratory time than field time, there is a component of laboratory time, roughly 20-30 minutes per sample, that is fixed regardless of sample size. This is the time taken to label bottles, wash and sieve the sample, and keep records (Ciborowski 1991). When processing time per sample begins to approach the fixed limit there is no advantage to further reductions in sampler size (see Sheldon 1984 for a detailed analysis). In addition, quality control procedures designed for large samples become much more laborious when replication increases, even if samples are smaller, because there are more taxon identifications and counts to be verified.

There are other practical considerations as well. If the sampler is too small, unpredictable error at the edges of the sampled area ("edge effects") become important, and a large number of zeros in species lists create havoc in statistical procedures (Resh and McElravy 1993). Finally, riffle samplers smaller than the average cobble on the stream bottom would be difficult to use in the field. The ideal size in most streams would still be much

smaller than conventional devices like the Surber sampler or Neill cylinder sampler. Sample sizes in the neighbourhood of 100 cm^2, one tenth the size of a Surber sampler, appear to work well (Mackie and Bailey 1981). For example, Scrimgeour et al. (1993) describe a sampler (based on Doeg and Lake (1981)) that samples benthic invertebrates on individual stones. In addition to the advantages of smaller samples and rapid sorting of replicates, the stone sampler reduces substratum heterogeneity, and samples a relevant unit of habitat for the animals being collected.

3.1.5 Persistence of Large Samples

The conclusion from this review is that in most ecosystems benthic invertebrate samplers should be as small as practically possible. But the advantage of smaller samplers is hardly a new finding. Well-used guides such as Elliott (1977) and Green (1979) have long recommended using the smallest practical sampler. Nevertheless, these recommendations have not translated into common practice. Said Resh (1979), referring to earlier work demonstrating the non-random distribution of benthic invertebrates:

> "These studies should have had a profound effect on the sampling design of benthic studies. However, this has not happened. From studies published in refereed journals to mimeographed reports of environmental impact statements, the same trend is apparent: quantitative studies are often based on very few benthic samples."

Yet there is little evidence of change in the succeeding 17 years. In Resh and McElravy's (1993) survey of published stream surveys, the Surber sampler or similar devices were still by far the most commonly used samplers, and 85% of the studies took five replicates or less. Choices of sampling gear and sampling intensity in environmental assessments seems to be guided more by tradition and convenience than by optimal design considerations (Ferraro et al. 1989). The challenge then, is not to just to provide better insights into sampling considerations, but to convince practitioners to adopt new approaches into their routines.

For one thing, smaller samples would require a change in the organization of sample sorting and taxonomic work. Biologists who specialize in benthic work almost universally charge by the sample, regardless of how large or small it is. A switch to larger numbers of much smaller samples would require that this whole pricing structure be re-thought, considering the fixed and variable time involved while allowing for unexpectedly dense or impoverished samples. As was concluded earlier in the discussion of sequential decision plans (Section 2.2), much closer collaboration between analysts and taxonomists will be necessary to improve cost efficiency in benthic surveys.

3.1.6 Optimal Sample Sizes for Site Comparisons

The calculations for determining optimal sample size apply to estimation of population densities for individual benthic species or the whole benthos at a single site. Ferraro et al. (1989, 1994) have extended this analysis to permit estimation of optimal sampling strategies for detection of a difference between sites, a more important objective in a biomonitoring study. Their procedure, based on the *t*-distribution, has three steps.

First, replicate samples are collected at a control (unimpaired) site, and at a site presumably suffering impairment of the magnitude the investigator wishes to detect. In their example, samples were taken in the marine benthos in Puget Sound, Washington, at a clean site and at a site near a fuel depot where petroleum contamination of the sediments was suspected. The mean difference in community response measures between the two sites, divided by the pooled standard deviation, produces the "effect size" of interest, i.e., the sampling program will be optimized to detect differences of that magnitude. The model can be based on any community parameter of interest, be it species richness, numbers of dominant or selected species, or any kind of compositional index.

Second, the total time or cost of each sampling scheme at each station is computed. In practice this means recording the time taken to sort and enumerate samples from each site and sampling device. Thirdly, a power analysis is conducted to determine the minimum number of samples needed to detect
the effect size chosen within acceptable limits of error, based on the *t*-test formula. The optimal sampling scheme will be that for which the product of required sample number times cost per sample is lowest.
This method is probably too expensive and too cumbersome for routine use in biomonitoring, but it has the attraction that any feature of the sampling program (sample size, replication, mesh size etc.) can be included in the analysis. Ferraro et al. (1989, 1994) tested their procedure at two sites in Puget Sound, and three sites off the coast of California. In both studies the most effective and efficient sampling program incorporated smaller sample sizes. In the California Bight, five replicate cores of 0.02 m^2 area could reliably distinguish control from degraded stations at less than one fourth the cost of five replicate 0.1-m^2 cores, the conventional protocol (Ferraro et al. 1994).

3.2 Mesh Size

3.2.1 Effect on Sampling Efficiency

The mesh size of nets used to trap benthos in the sampling devices like the Surber sampler, or of screens used to aid sample sorting in the laboratory, is of crucial importance to the effectiveness of the sampling method. Coarse-meshed nets allow smaller animals to pass, thereby biasing the sample and underestimating the real density of benthic organisms. Fine-meshed nets, on the other hand, trap detritus along with greater numbers of small organisms, resulting in samples that are time-consuming to sort and identify compared with those from larger-mesh nets (Environment Canada 1993).

By convention, macrobenthos has been defined as organisms that are retained by a 500 μm screen (Nalepa and Robertson 1981, Bachalet 1990), which corresponds roughly with those organisms that are easily visible to the naked eye. In fresh water this boundary includes the insects, oligochaete worms, molluscs, leeches, and macrocrustaceans such as amphipods, isopods and crayfish. It generally excludes microcrustaceans (copepods), flatworms, nematodes, rotifers and similar small-bodied organisms, which constitute the meiofauna. The problem is that, of the taxa considered part of the macro-invertebrates, not all members are large enough all the time to be trapped by a 0.5-mm net. Smaller species of mostly large-bodied orders, and especially early instars of aquatic insects and juvenile oligochaetes, tend to be under-represented in benthic invertebrate samples unless the capture net used is very fine.

The effect of mesh size on the retention efficiency of samplers has been a subject of research and debate for at least a half century (e.g., Jónasson 1958, Reish 1959). The uniform conclusion from many studies over the years is that coarse mesh nets lose smaller organisms and grossly underestimate the total density of macro-invertebrates at a site. For example, Kroger (1972) estimated the effectiveness of the Surber sampler by hand-picking animals from a dewatered river bed at points sampled with a 500 μm net a few hours earlier. He reported a mean of 4290 animals trapped in Surber samples, compared with a total density of 15 490. The missed animals were largely attributed to losses through the net, though spillage around the sampler also contributed.

Kroger's results are probably extreme, but other studies have confirmed the low retention efficiency of coarse nets for smaller organisms (see references in Resh 1979). Mundie (1971) showed that a 250-μm mesh net would pass about 90% of chironomid larvae and 50% of all other taxa from a small stream. Nalepa and Robertson (1981) compared retention of benthos samples from Lake Michigan using sieves of 595, 106 or 45

μm mesh. While virtually all the snails, fingernail clams (Sphaeriidae) leeches, amphipods and adults of the larger oligochaetes were retained by the 595 μm sieve, retention of immature tubificids, and smaller-bodied worms of the Naididae and Enchytraeidae ranged from 69% to as low as 2.5%. Essentially all the organisms were trapped on a 106-μm sieve.

Similar results have been reported in studies of invertebrate drift. Slack et al. (1991) found that a 425-μm mesh net passed half the Baetid and Ephemerellid mayflies, 71-94% of one family of stoneflies (Nemouridae) and >98% of the chironomid larvae. A 200-μm mesh trapped almost all organisms except the chironomids and a few earliest instars of other insects. The drift literature reviewed by Slack et al. (1991) and reproduced here as Table 4, illustrates that drift densities may vary by factors of one or two orders of magnitude according to the mesh size of the net used for sampling.

While very small members of any invertebrate group may be missed by a large-mesh sampling net, by far the most serious losses are among chironomid midges and immature oligochaete worms (Jónasson 1958, Kroger 1972, Storey and Pinder 1985, Slack et al. 1991). Retention efficiencies of 1-5% on a 500-μm mesh are not uncommon for these groups, especially for early instars or juveniles. More distressingly, retention varies widely even among species, depending on the diameter of the largest body part (Schlacher and Woolridge 1996), which for chironomids is the head capsule. Nalepa and Robertson (1981) showed that retention of chironomids on a 595-μm screen varied from 100% for large-bodied *Cryptochironomus* to as low as 21% for *Cladotanytarsus* (Table 5).

How fine must a net mesh be to approach 100% capture efficiency? The answer depends on the particular composition of the fauna at a given site, the season (and hence stage of development of larval insects) and whether the samples are sorted live. Retention of live animals, especially chironomids, by any mesh size is always less than for preserved animals because live animals actively wiggle through the net, while preservatives like formalin tend to render the specimen rigid. For most species, a net mesh in the range 200-250 μm is sufficient to catch all but the smallest members (Bachalet 1990, Slack et al. 1991). Chironomids are again exceptional, however. Retention of early instar chironomids on 200-μm mesh nets is often hardly better than on larger meshes, and mesh dimensions as fine as 100 μm or even less may be necessary to ensure adequate retention (Mundie 1971, Nalepa and Robertson 1981, Storey and Pinder 1985, Slack et al. 1991).

A biomonitoring program based on benthic invertebrates must face a compromise in the choice of mesh size. All but the finest nets will not suffice to capture small chironomids, and the sampling will inevitably be biased against smaller species and those represented by early instars. Estimates of species richness, evenness and

diversity will be similarly inaccurate (Bachalet 1990). On the other hand, an impressive phalanx of arguments can be marshalled against using fine mesh nets:

(1) Finer meshes provide better estimate of population densities, but also significantly increase the time, and hence the cost, required to process the samples (Schlacher and Wooldridge 1996). Given that most biomonitoring studies have limited budgets, increasing the sample processing time reduces the total number of sampling sites or the number of replicates at each site.

(2) The improved accuracy of finer meshes principally involves early instars of chironomids and other insects, along with immature oligochaete worms. These organisms are among the most difficult to identify, because of their small size and undeveloped features, and must frequently be lumped together at the family or order level. There is thus only a limited gain in information in return for a large increase in effort. It would be more practical to use that effort to collect more large-mesh samples whose members can be identified, and thereby derive better population estimates of larger organisms and later instars (Bunn 1995).

(3) The early-instar organisms better retained by fine mesh nets are not the best indicators of environmental conditions, because these organisms have not been in place long enough to respond to chronic conditions. Chironomids may have several to many generations in a summer, so early instars in a sample are likely to be only a few days to a few weeks old, and their numbers reflect more the fecundity of their parents than environmental conditions in the study reach. Site assessments can be confounded and rendered unnecessarily costly if the animals in the smallest size class are primarily ephemeral, patchily distributed juveniles (Ferraro et al. 1994).

(4) Smaller organisms retained by fine mesh nets make a negligible contribution to total benthos biomass (Nalepa and Robertson 1981) and stretch the definition of "macrobenthos".

The point of a biomonitoring program is to detect changes in environmental conditions through the response of the benthic invertebrate community. Hence, a bias in the sampling program can be tolerated as long as it is constant among sites and times, because differences in environmental conditions will always be determined comparatively. The extra information and improved accuracy of absolute population estimates that would be obtained from using fine mesh nets does not warrant the additional cost and expense of sorting and identifying the larger samples, not to mention the logistic difficulties of working with nets that rapidly clog with detritus. Hence, the ideal mesh size will be a compromise between sampling accuracy and practical limitations. Experience has demonstrated that mesh dimensions in the neighbourhood of 250 μm capture all but the smallest

instars of most organisms, with the exception of chironomids, but the nets needed to sample chironomids accurately are too fine to be practical for routine use. Therefore, 250 μm appears to be a reasonable choice of capture net mesh size, and is frequently recommended for biomonitoring in Canada (Anderson 1990, Environment Canada 1993).

3.2.2 Persistence of Mixed Mesh Sizes

It might be expected that, aquatic ecologists being aware of the exigencies of benthic invertebrate monitoring and the trade-off between mesh size and sampling effort, a consensus would have emerged on the mesh dimensions of nets for routine use. Unfortunately, this is not so. Resh and McElravy (1993) noted the mesh size of capture nets used in 84 published papers reporting on benthic invertebrate studies in the scientific literature, 44 from streams and rivers, and 40 from lakes and ponds (Table 6). In both sets of papers, mesh size ranged from <100 μm to >600 μm, with no clear standard emerging. Although 300-400 μm was the most commonly used size (27%) in flowing-waters, finer or coarser mesh sizes were nearly as popular. For comparison, an earlier literature review by Winterbourn (1985) reported that meshes in the range 200-300 μm were most commonly used, while the modal size of meshes used by respondents to a questionnaire (from the North American Benthological Society) was 590 μm (Resh et al. 1985). The larger mesh size in the last survey reflects the influence of the USEPA, which has adopted 590 μm as the standard mesh sizes for benthos surveys (Klemm et al. 1990).

A similarly even distribution of mesh sizes appears in the lake studies reviewed by Resh and McElravy (1993) (Table 6), except that limnologists tended to use finer mesh nets in their samplers, possibly because of the importance of smaller organisms in lake sediments (Nalepa and Robertson 1981). On the other hand, an earlier review by Downing (1984) found 450-600 μm was the most common mesh size in lake studies. More recent papers examined in the course of this review show the same range of mesh sizes in common use in both standing and flowing waters; larger meshes (500-600 μm) tend to be more common in qualitative surveys, such as those used for rapid bioassessment procedures. Hence, we cannot look to common practice to define an ideal capture net mesh size.

Of perhaps greater importance than the adoption of a uniform net mesh size is to encourage standardization of mesh sizes when different workers sample the same stream (Bunn 1995). Mining companies may hire external consultants to carry out routine biomonitoring, and the consultant doing the work frequently changes from year to year. Long-term records of improvement or deterioration in water quality that accumulate from these studies are confounded by changes in the size of the mesh used by different workers. Erman (1981)

compared results of four different baseline surveys done over a ten-year period on a shallow Colorado river. Although all the studies used Surber samplers, the reported fauna was very different (average similarity 34%), because of differences in mesh sizes, as well as taxonomy and laboratory procedures (discussed in Section 4). It is important to ensure comparability of the data that mesh sizes be similar from one round of monitoring to the next. Such a policy would mark a fundamental improvement in the quality of biomonitoring with a minimum of effort.

3.3 Sampler Bias

The choice of sampling device is a central issue in any environmental survey. No single sampler design is sufficient in all aquatic habitats, and the variety of aquatic habitats and situations to sample, along with continuing efforts by biologists to improve the accuracy, precision and convenience of sampling have lead to a dizzying variety of sampling devices. Only a small subset of these are used routinely, with the others relegated to experimental purposes or sampling difficult habitats (e.g., large stones, Doeg and Lake 1981; rock outcrops, Voshell et al. 1992; woody debris, Delong et al. 1993). An exhaustive review of all these samplers is not attempted here. A number of recent compendia compare and illustrate many samplers (Elliott and Tullett 1978, 1983, Merritt et al. 1984, Voshell et al. 1989); Merritt et al. (1984) provide an organized list of which samplers are best for which habitats, and Klemm et al. (1990) summarize the strengths and limitations of each, along with a comprehensive list of references.

The potential of sampling devices to cause bias in benthic invertebrate samples is a real concern in any biomonitoring program or aquatic ecological study. Bias refers to systematic error in the way samples represent the nature of the population or assemblage being sampled. Bias may be distinguished from variability, or sampling error, which pertains only to the variation in numbers from one sample to the next.

In the context of benthic invertebrate monitoring, bias may be of two kinds. A sampler may collect all or most of the species in the community but underestimate the actual numbers of each (*undersampling*). For example, a biased sample might contain 500 organisms of a given taxon (or as the sum of all taxa) when the real density in the stream is 1000 animals per sampled area. Or, a sampler might capture one species more readily than another, leading to a bias in the estimate of relative densities of the two species (*selective sampling*). As a consequence of selective sampling, the structure of the community may be misconstrued by underestimating the numerical importance of one group. At the extreme, total species richness may also be underestimated if some species are not sampled at all. Bias of this type often applies to whole groups of species that have a common feature such as small size, burrowing habit or cryptic appearance.

All sampling devices are biased to some degree. Undersampling is a universal problem, but selective sampling is frequent as well. Most of the literature examining this problem appeared before 1980 and is reviewed by Resh (1979). As a generalization, sampling bias arises from four factors:

- loss of organisms through the capture net (Section 3.2) in netted samplers, or around the sampler in backwash;
- loss of organisms through disturbance or turbulence when the sampler is set in place or strikes the surface, or through escape reactions of motile species;
- failure to remove all organisms from the substratum, especially those that cling to surfaces like rocks and leaves, and those in the deeper substratum, especially the hyporheos in streams; and
- inconsistency between operators collecting the samples.

In flowing waters, the bias associated with even the best samplers is considerable. For example, Growns (1990) demonstrated through repeated sampling of the same sites that only two-thirds of the organisms in a typical river bottom were removed by a pump sampler in the first sample. It took five repeated samplings to remove 98% of the organisms. In addition to undersampling, there was evident selective sampling as well; efficiency was best for epibenthic species of stoneflies (86%), dragonflies (83%) and true flies (Empididae) (82%), and worst for burrowing species of chironomids (66%) and caddisflies (59%).

These results are approximately typical of the kind and magnitude of bias often found in tests of stream samplers. Kroger (1972) estimated that barely one fourth of the insects in a mountain stream were captured in a Surber sample. The deep undersampling of chironomids by netted samplers, discussed earlier (Section 3.2) is a special instance of selective sampling bias. Naturally, because each sampling device is designed differently, the degree and type of bias they exhibit also varies. Comparisons among sampling devices repeatedly show that they do not collect individuals or species equally well (e.g., Boulton 1985, Robertson and Piwowar 1985, Storey and Pinder 1985, Wolcott et al. 1992, Brinkman and Duffy 1996). However, sampling devices deployed in identical habitats do generally capture the same set of common species, though the degree of undersampling may vary.

New sampler designs have addressed many of the problems in sampler bias listed earlier. Enclosed sampler such as the Hess, Box and Neill cylinder sampler overcome the problem of spillage and escapes in the Surber sampler (Klemm et al. 1990). Operator variance can be reduced through training (Clifford and Casey 1992) and by having one operator take all the samples. There has recently been a spate of interest in pump samplers, for collecting samples in variable currents, including slackwater areas where other net samplers will not function (Boulton 1985, Brown et al. 1987, Brooks 1994). It can be argued that organisms deep in the

hyporheos are not participating in the benthic community and would not be essential for water quality surveillance. Similarly, capture net losses are mostly of small organisms that may not be of great significance for biomonitoring (see Section 3.2).

Notwithstanding these improvements, some bias will always be present in any benthic invertebrate sample. Undersampling reduces the potential sensitivity of biomonitoring when it increases the variance of density estimates for sensitive species. Selective sampling is a problem when species that might have contributed to differentiating sites are excluded from the sample. Recall, however, that the goal of biomonitoring is to assess effects of pollution or disturbance, rather than to determine benthic population densities in absolute terms. Because changes in community structure at potentially disturbed sites are always determined comparatively, relative to the control sites, some bias in the sample can be tolerated, as long as two conditions are met: (1) the bias is small enough that most of the benthic community, and most of the organisms sensitive to the disturbance, are included in the sample; and (2) the bias is equal at all sites. If a sampler badly undersamples a taxon that is very sensitive to a mine effluent, effects of the effluent on downstream sites may pass undetected. On the other hand, if a species is undersampled much less at one site than another, the effect of the effluent can be exaggerated or underestimated.

To address the first requirement, Long and Wang (1994) have proffered a method for comparing the capture efficiency (undersampling bias) of two sampling devices, based on the ratio of mean to standard deviation, rather than the absolute number of animals caught. If the difference in mean/SD ratios between two sampling devices was 0.2, then number of organisms in a sample from the first sampler would be within the 95% confidence limits of the other sampler 95% of the time, which is tantamount to saying there is no practical difference between them. A differences in mean/SD ratios of 0.3 is considered moderate and 0.4, large (Long and Wang 1994). A paired-sample *t*-test can be used to compare the samplers. This approach does provide a way to compare samplers for equal bias, although in practice a large number of samples is needed for a test of reasonably high power.

Serious undersampling of sensitive species is not generally a problem if modern equipment and procedures are used. Even in the presence of bias, samples collected in similar habitats with the same sampling device by the same experienced operator will adequately reflect the composition and numbers of common, larger invertebrates at the site. These species contribute the most to detecting and understanding pollution effects, because they are easily identified, abundant, and exposed to environmental conditions for some time. Therefore, the bias inevitable in these samples is not a major concern for the efficiency of biomonitoring. In fact, the bias toward

larger, epifaunal species may improve resolution by excluding species less likely to show subtle responses to surface water quality.

The greater concern for biomonitoring, and the issue that is more amenable to solution, is the problem of unequal bias among sites. A difference in bias is most likely to appear between sites if sampling locations are not chosen carefully or if identical habitat (riffles) do not occur at all sites. A difference in bias between years may arise from a change in sampling device. Samples from the same location taken in different years by different people, even when using the same sampler, are often difficult to reconcile (Erman 1981). Hence, errors induced by sampler bias are most effectively controlled by adoption of standard sampling methods, insofar as that is possible, and by careful selection of sampling locations.

4. Laboratory Methods

4.1 Sample Sorting

Aquatic biologists agree that the most tedious and time-consuming task in a benthic invertebrate survey is separating the organisms in a sample from the sediments and organic detritus (Mason and Yevich 1967, Ciborowski 1991, Wilhelm and Hiebert 1996, Brinkman and Duffy 1996). Sorting time is a stubborn hurdle limiting the efficiency of benthic sampling. Over the years many innovations have been suggested to help reduce sorting time, and new ideas are constantly being tried. A brief summary of the more successful methods is presented here. Magdych (1981), Rossillon (1987) and Meyer (1990) provide brief reviews of facilitation methods and entries to the literature.

Methods to reduce sample sorting time can be subdivided into two classes: facilitation and subsampling. Facilitation refers to methods that speed separation of animals from debris and sediments, while subsampling refers to sorting only a portion or portions of the whole sample. Most sorting methods can be used either on a whole sample or on a subsample, and the most effective protocols often combine selective subsampling and facilitation methods. As with every other aspect of a sampling program, the utility of a sorting method is determined by efficiency, defined as the percentage of the total number of invertebrates removed from the sample, and time required, which determines cost. Avoiding bias is also an important consideration.

4.1.1 Facilitation

Four kinds of facilitation methods are in general use (Barmuta 1984): sieves, elutriation, dyes and flotation. Each of these methods has many variants, and different methods may be more or less advantageous under different circumstances. A perfect facilitation method, one that separates all the organisms in any kind of sample without bias and relieves the investigator of the tedium of hand sorting, has yet to be found. The four methods are described next, in order of popularity.

1. Sieves

Sieving samples is probably the most widely used facilitation method. The procedure is to screen the sample through a series of two or three (rarely more) sieves of decreasing mesh size so that the material in the sample, invertebrates and detritus, is separated into a set of size-based fractions. Fine particles, silts and clay are removed from the sample, making the method particularly attractive for samples from depositional areas with

a fine substratum. Sieving can also be used to wash the sample free of preservative, especially formalin (Environment Canada 1993).

The mesh sizes used depend on the nature of the sample and the preferences of the investigator. The smallest sieve in the series should be the same size as the net mesh dimensions of the sampler used in the field (Anderson 1990). A series like 4, 2, 1, 0.5, 0.25 mm is more or less typical (Ciborowski 1991), but most workers do not use that many sieves (e.g., Rossillon 1987). Many consultants use only one sieve, to separate a coarse and fine fraction, or sometimes two sieves if the fine fraction needs to be subdivided further. Naturally, the sieve sizes used in any particular study can be varied to suit the nature of the samples.

Sieving samples helps with subsequent sorting by separating the sample into classes containing a uniform size of particles, both benthos and detritus. The largest sieve traps the large sticks, leaves and stones, while the smallest will contain only sand and fine organic detritus. The larger organisms are removed in the coarser sieves, and it is much easier to pick out small organisms like chironomids when coarse detritus has been removed. Sieving is less useful in samples with large amounts of detritus that can clog the screens. Filamentous algae in particular are a nuisance for sieving. In addition, there is a risk of mechanical damage to fragile organisms, especially mayflies and oligochaete worms, that can ruin the specimens or damage body parts necessary for identification (Resh 1979). In discussions, consultants disagreed about whether this is a serious problem, with some maintaining that proper field preservation will prevent breakage.

An interesting variation on the normal sieving procedure is described by Wilhelm and Hiebert (1996) who used large screens of 500 or 275 μm mesh mounted in bottomless buckets to filter the samples. The screens were then slowly immersed in water, and the animals were trapped in the surface film, from which they were easily skimmed off. Efficiency of removal ranged 39-92% (mean 74%) for benthic samples from a small stream, but the time savings was not great because the residuum must still be searched for the remaining animals. A substantial time savings might be realized at high population densities, however.

2. Elutriation

Elutriators separate organisms in a sample from debris and sediments by agitating with water or air. There are many designs, but the model described by Magdych (1981) is typical. It consists of a long tube with a sealed opening at the bottom through which a water current can be introduced, and an overflow spout at the top, leading to a sieve. The sample is placed at the bottom of the water-filled container, and the tap controlling

water flow up from the bottom is turned on. The current agitates the sample, separating light animals from detritus, and carrying them upward, where they are carried into the overflow spout and trapped on the sieve.

Elutriators work best with samples that are heavy with gravel and inorganic sediments. In samples of that kind Magdych (1981) reported 96 ± 3.5% efficiency of removal of benthic animals, but other workers report lower, and variable, efficiencies. The principle limitations of elutriation are: (1) it is biased against heavy-bodied organisms, especially molluscs and stone-cased caddisflies that are not carried upward like lighter organisms; (2) light organic detritus will be flushed out with the animals; and (3) it may take a long time to process a single sample. Brinkman and Duffy (1996) elutriated wetland core samples for an hour apiece; mean recovery in five replicates was 69.3 ± 26.1%.

3. Dyes

Selective dyes that stain organisms a conspicuous colour improve sorting efficiency by making individual animals easier to see against the background of detritus or sediments (Lackey and May 1971, Williams and Williams 1974). Rose bengal, which stains animals pink, is the most commonly used dye (Resh and McElravy 1993), but Phloxine B (Mason and Yevich 1967) and Congo Red (Brinkman and Duffy 1996) have also been suggested. The last authors also tested Rhodamine B combined with sorting under ultraviolet light (to make the animals fluorescent) but found there was no improvement over conventional dyes.

Dyes can be added to the samples in the laboratory or mixed with the preservative and added to the samples immediately in the field (Klemm et al. 1990). The latter method is more popular because rose bengal, for example, requires 24 h for complete penetration of the stain (Anderson 1990). While dye-staining samples has been shown to improve efficiency of benthos recovery from samples (Mason and Yevich 1967), use of this technique varies widely, mostly according to individual preferences (Environment Canada 1993). Some researchers and consultants insist that dye-staining sharply improves recovery, while others maintain that the benefits are minimal (Cromar and Williams 1991), and the dyes may interfere with identifications. Rossillon (1987) found that the improved efficiency from addition of rose bengal was minor compared with that from other facilitations. Even where dyes do not result in time savings, they may still improve accuracy because fewer animals are missed (Resh and McElravy 1993).

4. Flotation

Flotation is perhaps the oldest method of separating animals from detritus (Anderson 1959) and the method that has produced the most variants. Inorganic particles, and most organic material other than fresh leaves and algae, have a specific gravity >1.12, while the specific gravity of aquatic organisms is less. Hence, when a sample is placed in a solution of sugar or other solutes with a specific gravity above 1.12, the detritus will sink to the bottom while the animals will rise to the top, where they can be skimmed off. A dense sugar solution (about 300 g/L) is widely recommended for flotation. Many other solutes have been tested, including magnesium sulphate, D-mannitol, calcium chloride and sodium chloride (Klemm et al. 1990) but sugar is preferred because it is cheap, readily available, nontoxic and uncharged in solution. Formal methods of flotation involve adding the sample to a beaker or column of solution and removing the floating organisms after the sample has separated, but sorting can also be facilitated by adding a few tablespoons of sugar to formalin-preserved samples in the sorting pan and stirring gently to separate animals from detritus (Klemm et al. 1990).

Flotation methods are not without drawbacks. The main limitations are these:

(1) Organisms in a hypertonic solution lose water and eventually sink again when their specific gravity matches that of the solution (Cromar and Williams 1991);

(2) Separation from detritus is imperfect. Some organic matter, especially fresh litter and small particles, floats along with the animals. Conversely animals entangled in moss or algae may not float. The method works best on samples with mostly sand or inorganic debris (Mason and Yevich 1967);

(3) Most importantly, the method is strongly biased against denser organisms. For example, Rossillon (1987) reported 100% separation of insects from detritus, but <30% for molluscs and flatworms. Sand-cased caddisflies, such as the widespread *Helicopsyche borealis* will also be undersampled by flotation (Resh 1979).

Many or most of the organisms in the sample may be removed through flotation. The remaining detritus must then be examined for clams, snails and other heavy organisms left behind. Flotation methods work better when the extraction is repeated (Rossillon 1987, Anderson 1990) but of course that adds more time to sample processing.

A promising extension of the flotation method, especially for samples rich in fine organic detritus, is to combine it with centrifugation (Cromar and Williams 1991). In this method the sample is immersed in a denser sugar solution (600 g/L) with a specific gravity of 1.17. The sample is then centrifuged for about 45 s to speed separation; mineral particles sink to the bottom, fine organic detritus is thrown part way down, and organisms remain near the top. The inventors claim that in five samples rich in organic matter, mean sorting time was

reduced from 302 ± 71 min to 73 ± 9 min with better recovery of organisms (374 ± 81 per sample versus 226 ± 48) and no apparent bias. One would expect that the mineral layer would still need to be examined for molluscs, however.

A similar idea from Barmuta (1984) uses phase separation in a mixture of kerosene and alcohol/water. When the sample is agitated and allowed to settle, organic detritus migrates to the alcohol phase, but animals tend to concentrate at the interface. The method showed promising recovery efficiency (88% for one extraction, 95% if repeated) but did not work for crustacea. Other methods might be preferred that do not require flammable chemicals.

In evaluating any facilitation method for routine biomonitoring, the effect of the method on quality assurance targets must also be considered. Programs such as the environmental effects monitoring for pulp mills specify 95% recovery of the animals in every sample, as determined by random re-sorts, and competent commercial laboratories maintain internal quality checks to ensure these minimum standards are met. A facilitation method that saves time but leads to lower recovery efficiencies would not be acceptable for routine use.

All facilitation methods have limitations, but these short-cuts can produce substantial time savings in benthic invertebrate studies. A survey of researchers and consultants working in benthic invertebrate ecology suggests that sieves, stains, flotation and elutriation can reduce time taken to sort benthic samples by 15% to as much as 45% (Table 7). Of course, the methods are not mutually exclusive and the most efficient protocols combine elements of several methods. Sample sorting is usually the single most time-consuming step in a benthic sampling program (Sheldon 1984), so any innovation that saves time without causing imprecision or bias or lowering recovery efficiencies should be embraced. From the reverse angle, improvement in sorting efficiency would allow inclusion of more samples in a benthic study for the same amount of money (Resh and McElravy 1993).

4.1.2 Subsampling

Subsampling is a special case of sorting facilitation so it will be discussed separately, and briefly, here. The large literature on subsampling and the vast array of devices that have been devised for subsampling benthic invertebrate samples lie beyond the scope of this report (see Hickley 1975, Wrona et al. 1982 and Sebastien et al. 1988 for overviews and apparatus). Most workers are aware of the importance of minimizing bias when subsampling benthic samples, and of taking as large a subsample as possible.

The problem inherent in subsampling is that it reduces sample size and thereby reduces the potential precision of estimated population densities. A small subsample still estimates the total number of organisms in the whole sample, but the uncertainty about that estimate becomes larger as the subsample gets smaller (Sell and Evans 1982). Presumably, the high number of animals in the sample was what prompted the decision to subsample, so for the common species the loss of precision is not an issue. For less common species, however, accuracy of the population estimate may be compromised by the reduced size of the sample (Meyer 1990). Many workers have suggested, based on the Poisson distribution, that reasonable accuracy (often taken to be a standard error 20% of the mean) can be obtained when at least 100-200 animals are contained in the subsample (Lund et al. 1958, Hickley 1975, Elliott 1977, Sell and Evans 1982, Rossillon 1987, Klemm et al. 1990). That rule will work for total numbers of all species, or the dominant species, but variances for less common species will be larger (Wrona et al 1982).

A number of solutions to these related problems have been offered. The most elegant solution is to fraction the sample into equal size fractions using sieves, and then subsample only the fraction or fractions containing too many animals (Reger et al. 1982, Meyer 1990). All the organisms in the other fractions would be sorted, so for species confined to those fractions no precision would be lost through subsampling. This approach has the added advantage of producing a uniform size distribution of the detritus and animals, which makes random sampling easier to attain (Anderson 1990).

If population estimates of less common species in the abundant size fraction were deemed necessary, sorting could continue until the counts for these species exceeded 100. If the first few subsamples provided high enough counts of the common species, they would be ignored in subsequent subsamples (Wrona et al. 1982). The alternative is to decide on a fixed number of subsamples, and accept the higher variances of the less common species. Where detection of community-level effects of disturbance or pollution is the intent, continued counting of rare species defeats the purpose of subsampling, namely to save time and effort. Fewer than 100 animals in a sample, down to as few as 20, will still be enough to produce a density estimate with ± 50% precision (Figure 5), sufficient to detect many site differences (Wrona et al. 1982). Moreover, there will always be some species for which the precision of the estimate will remain poor even if substantial extra effort is expended to sort more subsamples, up to and including sorting the entire sample (Sebastien et al. 1988).

Subsampling will also affect the estimate of number of species in the sample. Species that are represented by a few individuals in the full sample will have the least accurate population estimates in the subsample, and by chance may be excluded completely. Total number of taxa (species richness) is affected by subsampling, often

in unpredictable ways, and the richness of the full sample cannot be back-calculated from the subsample the way species abundances can be (Environment Canada 1993).

The most reliable way to estimate the number of taxa in a sample from subsamples is to serially sort a number of small subsamples and plot number of taxa encountered against sorting effort, in effect creating a species-area curve. Subsampling may end when the curve approaches an asymptote (Courtmanch 1996). In practice, the curve may not become asymptotic until >50% of the sample has been sorted, effectively negating the intent of subsampling. However, if subsamples have been taken from many replicates, these can be plotted cumulatively in the same way, and if an asymptote is approached a reasonable estimate of species richness for the site may be had, albeit without confidence limits. Vinson and Hawkins (1996) recommend two-phase sampling to estimate species richness, first covering the whole sample looking for large, rare organisms (like Perlid stoneflies) and then subsampling the remaining fraction. The same end will be achieved by sieving and subsampling only the fine fraction, as described earlier.

The best solution to the subsampling dilemma is to take smaller samples to begin with and avoid subsampling altogether. Section 3.2 argues that most benthic samplers for flowing waters take samples that are much too large and cost-efficiency could be greatly improved if a larger number of smaller samples were collected. Small samples can be processed quickly in the laboratory and would seldom require subsampling. Rare species are less reliably estimated by smaller samples, but the contribution of these species to detecting environmental stress on benthic communities is minimal. Section 4.3 presents a case for deleting statistically rare species from analysis; there would thus be no loss, and a slight gain in time saved, if these species were not collected in the first place.

4.2 Taxonomic Resolution

Taxonomic resolution refers to the exactness of identifications attached to the organisms collected in a biomonitoring sample; it is also referred to as *taxonomic penetration* (Cranston 1990). Complete taxonomic resolution of a sample would be to identify all the organisms in it to species; but in practice some or all members of the sample might be identified to genus, family or higher taxa. Before examining the lively debate on the effect of taxonomic resolution on benthic invertebrate monitoring, a semantic detail must be settled. The literature on biomonitoring refers to "higher" or "lower" levels of taxonomy, but there is no unanimity as to what the terms mean. Here, species is the *lowest* level of taxonomy, and genus, family, and order represent *higher* taxonomic levels. The reader is warned, however, that in some literature the terms may mean exactly the opposite.

The effect of taxonomic level on the sensitivity of biomonitoring programs has been debated for some time, yet there is still no broad consensus among researchers. For example, Resh and McElravy (1993) surveyed 31 recent papers that touched on the issue and found that 18 emphasized species-level identifications, nine recommended using higher taxa under some circumstances and four suggested using both, depending on the objectives of the study. Fortunately, recent work has concentrated on examining the quantitative effect of taxonomic resolution on results of biomonitoring studies, which has allowed a more objective analysis of the problem. It is apparent now that the need for specific identifications depends on both the spatial scale of the study and the sensitivity required (Herricks and Cairns 1982).

4.2.1 Value of Species-Level Identifications

The first, and most powerful, argument for specific identifications is that species are a basic unit of biological organization, and because each species is unique, it has important attributes -- life-cycle, habitat, sensitivity to different kinds of pollution -- that are not shared by any other species. A species list from a given site will thus always contain the greatest amount of biological information compared with higher taxonomic levels (Resh and Unzicker 1975). A site assessment based on species can take full advantage of ecological research on populations of individual species, or comparisons among closely-related species. This information is suppressed when species are lumped together into genera or higher taxa.

Among taxonomic categories, attributes like sensitivity to copper toxicity or tolerance to sedimentation can only properly be assigned to species, which by definition are groups of genetically similar organisms. Values for such attributes for higher taxa are means of the values for all the component species. The resolution of specific identifications are lost at the genus level because high or low values of the attribute in question possessed by different species cancel out in the average. Resh and Unzicker (1975) illustrated this point with a standard table (from Weber 1973) listing pollution tolerance categories for 61 species of freshwater insects: where tolerance classes had been established for species, genera tend to fall into two or three classes, because they have tolerant, facultative and intolerant member species. Thus, while an individual species may have quite narrow ecological limits and pollution tolerances, the genus will be found over a wider range of conditions (represented by different species at each location), thereby reducing the sensitivity of biomonitoring indices based on genera.

As an example, Table 8 presents presence/absence data on 57 species of Chironomidae from a second-order, limestone stream in Ohio receiving a complex heavy-metal effluent from a metal-plating industry (Waterhouse and Farrell 1985). Copper concentrations, an indicator of the level of metal contamination, declined from 336

µg/L at station 1 to 74 µg/L at station 5. Chironomid species sort themselves along the metal-contamination gradient according to the tolerances of individual species, and there are marked differences among them, even within a single genus. For example, *Polypedilum convictum* is among the most tolerant species, occurring at all five stations, but both *P. halterale* and *P. scalaenum* are found only at the cleanest site. An assessment based on genera would have marked *Polypedilum* at all five sites and missed an important indicator of the contamination gradient. Similar variations among species occur within most of the other genera observed. Even *Micropsectra*, represented by only two species, apparently contains one tolerant and one intolerant species (Table 8).

Two counter-arguments can be raised against specific identifications. First, while every species is slightly different in its environmental requirements, the hierarchial structure of classification guarantees redundancy in the information content at specific, generic, or higher taxonomic levels (Ferraro and Cole 1995). Redundancy is complete in monospecific phyla, but there is considerable redundancy even in large genera because proper taxonomy groups species according to their relatedness.

In the data of Waterhouse and Farrell (1985), 13 of 27 genera are represented by a single species (Table 8), and generic identifications were sufficient to detect the contamination gradient. These authors credited the agreement between species-level and genus-level analyses to the ability of a robust species distribution to withstand a certain level of information loss when grouped into genera, rather than any similarity of response among closely related species. In other words, generic identifications merely diluted the pattern shown by the metal-sensitive species with "noise" from other species that did not respond strongly to the gradient. The degree of information loss increased with the size of the genus. Genera with many species did not contribute much to the differentiation of stations compared with genera of one or two species because at least one member of the large genera was bound to be present at every station. Higher taxa will show the same effect, with progressive loss of information at each level.

The redundancy argument applies at any level of taxonomy. In both marine and freshwater ecosystems, it has been repeatedly demonstrated that samples identified to the genus, family, or even order level are sufficient to detect strong gradients of pollution, or to discriminate clean sites from affected ones (see later). The success of high-level taxonomy at detecting disturbance gradients, in spite of the known variation in tolerance among species in a taxon, evidently arises because variation in tolerance within any given genus, family or order is still much less than differences between them (Wright et al. 1995).

Closely related species are placed within a genus, and closely related genera share the same family. It follows that environmental requirements and tolerances will be broadly similar within any group, with the degree of

differentiation weakening at successively higher levels (Marchant et al. 1995). Mason et al. (1985) assigned 172 benthic invertebrate taxa from an Ohio river to one of 10 categories of pollution tolerance. The eight species within the midge genus *Cricotopus* occupied only two categories; the subfamily Orthocladiinae, of which *Cricotopus* is a member, spanned five categories; and the family Chironomidae was represented in all ten categories.

But even at the order level, different groups of insects (Ephemeroptera, Plecoptera, Trichoptera, Diptera, etc.) are well known to have broadly consistent sensitivities at least to organic pollution; for example, stoneflies and mayflies are usually the first species to disappear at enriched sites, and this sensitivity is the foundation of quick-assessment procedures based on the number of families of Ephemeroptera, Plecoptera and Trichoptera present at a site. Using the classifications of Mason et al. (1985) just the number of families in these three taxa would be sufficient to differentiate broad levels of impairment:

Tolerance Category	Number of Families Represented
1	11
2	11
3	8
4	4
5	4
6	1
7-10	0

The second argument against specific identifications suggests that individual species are *too* sensitive to environmental change; that is, they will respond to minor changes in environmental conditions from one site to the next unrelated to any pollution or disturbance gradient, and hence obscure the analysis (Warwick 1988, Smith and Simpson 1993). Benthic invertebrate species are very closely attuned to their physical habitat, and small changes in water depth, current velocity or substratum can lead to replacements of one species by another at a particular micro-site. This responsiveness is the major source of background variation in the density and community composition from one place to another along any water course. If there is substantial habitat variability between stations, species data may introduce "noise" and actually reduce the sensitivity of the analysis, while higher taxa would respond less to fine-scale habitat differences, (reduce noise) and let the pollution signal penetrate (Vanderklift et al 1996).

To return once again to the data of Waterhouse and Farrell (1985), it is evident that the distribution of some species is governed by factors other than the metal gradient. If *Pentaneura currani*, *P. fimbriata*, and *P. pilosella* were prevented only by metal toxicity from occurring at sites upstream from Station 4, they should logically be present at Station 5 as well, but they are not (Table 8). Of course, competition with other species is itself a factor influencing distribution, and these species may be competitively excluded from the cleanest site. Nevertheless, the distribution patterns of many species in Table 8 are too irregular to be attributed solely to the contamination gradient. In marine studies, it has been argued that higher taxa are better for detection of strong pollution gradients because they suppress the individual variation in site preference among species (Warwick 1988, Smith and Simpson 1993). However, Wright et al. (1995) showed that species-level taxonomy always gave at least marginally better discrimination of clean from polluted sites in an Australian river.

These conflicting results arise because of the confounding effects of scale and severity of pollution with taxonomic effects. Genera and higher levels, being composed of what we believe to be closely related species, tend to be similar in their requirements for large-scale habitat characteristics, while species differences occur more at a microhabitat scale (Green 1979, Waterhouse and Farrell 1985). For example, Wiggins and Mackay (1978) could place most genera of Trichoptera, Ephemeroptera and Plecoptera in ranges along the continuum from headwater streams to rivers, according to the ecological requirements, mostly for food resources, temperature and water velocity, of each. The distribution for the caddisfly family Polycentropodidae is shown here (Figure 6) as an example. Individual species within each genus would be distributed among habitats (usually overlapping) within the generic range according to the narrower environmental demands of each. Hence, genera or highe levels are effective indicators of broad-scale differences between sites, while species within genera respond to finer differences.

This effect of spatial scale explains why genus, family and even order may often be sufficient to distinguish environmental quality among sites covering a broad geographical area, or in which changes in habitat are relatively large. Magdych (1984) found firm relationships between the distribution of mayfly genera and physical-chemical variables in a stream with discharge, salinity and food-supply gradients. Family-level data effectively described longitudinal trends in water quality along a 500-km reach of the French Rhône River (Bournaud et al. 1996). Individual species vary in their sensitivity to pH, but regional effects of pH are apparent at the genus level (Hall and Ide 1987).

Similarly, even higher taxonomic levels will suffice to elucidate the effects of severe disturbance or a steep gradient of pollution. These kinds of disruptions of aquatic habitats generally have conspicuous manifestations like reduction in total species richness, reduction or increase in population density, and disappearance of entire

high-level taxa (e.g., mayflies) or feeding groups (e.g., filter-feeders). Hence, as mentioned earlier (Section 1.1), no sophisticated methods or detailed taxonomy are necessary to demonstrate the effect of severe disturbances (Gray et al. 1990, Resh and McElravy 1993). But subtle, small differences among sites are better resolved when the taxonomy is taken to the lowest level possible.

4.2.2 Statistical Considerations

The taxonomic resolution required also depends strongly on the nature of the analysis. Strictly statistical approaches that work with densities and numbers of species without distinguishing among them, tend to be rather insensitive to taxonomic level (Resh and McElravy 1993). The effect of taxonomic resolution on multivariate methods, in particular ordinations, has been studied most. In marine systems, R.M. Warwick and his co-workers have demonstrated repeatedly that broad-scale patterns of pollution from oil exploration and similar activities can be detected on ordination plots (multidimensional scaling) as effectively with data at order, class or even phylum level as with specific data (Heip et al. 1988, Gray et al. 1988, 1990, Warwick 1988, Warwick and Clarke 1991, 1993, Agard et al. 1993). Vanderklift et al. (1996) confirmed the observational conclusions of Warwick with a more quantitative analysis based on metal contamination around a lead smelter. These results must be extrapolated with care to fresh waters because of the greater phylogenetic diversity in marine systems (Gray et al. 1990). Benthos of lotic fresh waters tends to be strongly dominated by one class, the Insecta.

In freshwater ecosystems, ordinations are generally more effective when species-level data are used as input (Marchant 1990, Furse et al. 1984, Wright et al. 1995). However, in most cases the loss of sensitivity from generic or even family-level identifications is not large (Faith et al. 1995, Wright et al. 1995, Furse et al. 1984). Lower levels of taxonomy are relatively more important at small spatial scales than at large ones, such as comparisons among rivers (Armitage et al. 1987, Marchant et al. 1995, Bournaud et al. 1996). Similarity indices alone generally show the same patterns at higher taxonomic levels as with species data, although again there is a slight loss of resolution when species data are collapsed (Faith et al. 1995). Waterhouse and Farrell (1985) found that a variety of presence/absence similarity indices showed the same pattern among metal-contaminated stream sites whether calculated from species or genera of Chironomidae.

Incomplete taxonomy will severely underestimate the true species richness of an ecosystem or study site (Resh and Unzicker 1975, Harper and Cloutier 1986, Cranston 1990). For a biomonitoring program, however, this is only a problem if the error is unequal between sites. If not, then comparisons of changes in species richness between sites or over time should still be valid, because the bias will be the same at all sites. There will

inevitably be some loss of resolution, however, from higher taxonomic levels, because a given family or genus could still be represented at an affected site even if three of its four member species had disappeared.

The Shannon-Weaver diversity index is sensitive to the number of "species" involved in its calculation, and therefore declines with higher levels of taxonomy. As with species richness, the real diversity of a site can be badly underestimated when identifications are not done to species (Hughes 1978). Nevertheless, conspicuous differences in diversity between sites based on species-level identifications are preserved when data are lumped into genera or families (Hellawell 1977, Bournaud et al. 1996). Osborne et al. (1980) found that diversity indexes calculated at the genus or even family level were sufficient to detect a strong gradient of disturbance associated with mining.

From a strictly pragmatic viewpoint, pooling lower taxa into larger groups may contribute to the strength and ease of statistical analyses. First, the variance of pooled taxa will often be improved because of the increased sample size (Keough and Quinn 1991). Second, pooling taxa tends to remove zero density estimates, which complicate multivariate ordinations. Third, pooling the lower taxa increases the ratio of sites to taxa used in ordination and multivariate analysis of variance. Where the number of taxa exceeds the number of sites, invariably the case when specific identifications are used, results from ordinations may not be stable, and MANOVA cannot be used (Norris and Georges 1993).

Conversely, analyses that depend on ecological information about the species found at each site are decidedly more powerful when organisms are identified to species. These studies can take advantage of information on the biology and tolerances of different species to look more deeply into the nature and causes of the disturbance. For example, construction of a bridge across a small stream in southern Ontario caused both sedimentation and organic enrichment (from mulching for revegetation) in the reach downstream (Taylor and Roff 1986). Large increases in net-spinning caddisfly populations downstream were dominated in the first two years after construction by one species, *Hydropsyche slossonae*, which could tolerate siltation and take advantage of the enhanced food supply. Less silt-tolerant species, *H. sparna* and *H. betteni,* became dominant later when silt was flushed out of the system. The specific identifications in this study permitted separation of the two influences on the system, that would not have been possible with less detailed taxonomy.

Again, however, genus-level identification can still provide much useful information about freshwater benthic communities when combined with ecological data about the habits and habitats of the genus. The functional group classification of Merritt and Cummins (1984) assumes that genera of aquatic insects can be classified according to trophic relationships (predator, shredder, filterer etc.), habits (clinging, burrowing, climbing,

swimming) or habitats (lotic, lentic, erosional, depositional) and that this classification will be more or less consistent for all species within the genus. This classification builds on the idea of Wiggins and Mackay (1978) that insect genera represent a sort of ecological type, of which species present minor variations. Rooke and Mackie (1982) used this idea to develop a method of invertebrate monitoring based on the habitats and habits of genera (and sometimes higher levels of non-insect taxa) and showed that it could reveal both the degree and the nature of changes caused by an impoundment.

Biotic indices, which are attempts to compress information about invertebrate community response to pollution into a single number, are most accurate when they are based on species. This conclusion follows immediately from the observations discussed earlier that environmental tolerances are attributes of individual species, not higher taxa. Nevertheless, many biotic indices and similar scoring systems for quickly evaluating a site are based on genus-level or even family-level identifications and work reasonably well. (Hilsenhoff 1987, Metcalfe 1989, Chessman 1995). Again, however, these systems are most effective for broad-scale comparisons of sites or for detection of large disturbances (Chessman 1995). Recently, Hilsenhoff (1988) has developed a very fast, simple version of his genus-based biotic index that uses field identification of insect families. As would be expected, it is less sensitive still than the original index, but can still identify, and tentatively rank, streams suffering organic loading.

4.2.3 Limitations of Taxonomy

The strongest argument against complete taxonomic resolution is pragmatic: identification of many species of freshwater invertebrates is difficult or impossible with present knowledge. Most larval insects (and adults, in the case of beetles) found in fresh waters can be confidently identified to genus by any competent biologist armed with up-to-date keys. Identification of larval midges of the ubiquitous family Chironomidae is rather more difficult; generic identifications in this family require clearing and mounting head capsules for examination of mouthparts under a compound microscope, an exacting and time-consuming exercise.

Most aquatic insects cannot be identified to species without rearing the immature form to the adult (Merritt et al. 1984); that is how Waterhouse and Farrell (1985) obtained the list of chironomid species in Table 8. Such work is far too demanding and too time-consuming to be practically applied to the large number of species collected in a biomonitoring study. Species-level identifications are possible for some immature insects with conspicuous markings, but these species are not the majority. Hence, for all practical purposes the genus is the lowest level of taxonomy for immature aquatic insects.

A number of other groups pose similar problems. Many oligochaete worms can be identified to species only by examining setae and internal reproductive organs on mounted specimens. Immatures cannot be assigned to species. Water mites, some molluscs and most other minor members of the benthos require a great deal of specialized expertise to arrive at specific identifications. Finally, any sample of the benthos will contain a variable number of early-instar insect larvae that are too incompletely developed to permit complete identification.

The result of all these practical impediments is that identification of benthos samples is inevitably incomplete. Out of more than 90 000 organisms collected in a study of the Rhône river, only about 26 000 could be identified to species (Bournaud et al. 1996). Scientists and consultants involved in aquatic biomonitoring generally seek identifications to the "lowest practical level" (Anderson 1990, Klemm et al. 1990). It must be remembered as well that the taxonomy of even the well-studied insect groups is incomplete, and subject to periodic revision. For example, a cluster of recent studies has substantially revised the taxonomy of the common mayfly family Baetidae, including re-assignments of species to genera and descriptions of new genera and species (Allen 1984, Pescador 1985, Waltz et al. 1985, Waltz and Mccafferty 1987a,b,c, Provonsha 1990, Mccafferty and Waltz 1990, Mccafferty 1992).

Resh and McElravy (1993) point out that taxonomic obstacles can be at least partially overcome by exchange of information among workers and especially closer co-operation with specialists in the various invertebrate groups. Another remedy is to separate putatively different species (or higher taxa) but not attempt to assign them names (Cranston 1990). However, the additional effort required to drive taxonomy to the species level, especially for difficult, speciose, or abundant taxa, can sharply increase the time and cost of a biomonitoring study. It has been argued that the cost of species-level identifications is minor once personnel have been trained and know which species to expect (Lenat and Penrose 1980). Others, however, have found that there is a continuing cost associated with complete taxonomic resolution (Kaesler and Herricks 1980, Furse et al. 1984). Many workers question whether that cost is worth the return in terms of increased sensitivity (Warwick 1993, Warwick and Clarke 1993, Ferraro and Cole 1995, Vanderklift et al. 1996).

The unavoidable trade-off between taxonomic penetration (information) and cost in biomonitoring returns us again to the question of optimal allocation of effort. If generic or higher-level taxa provide sufficient information to identify environmental degradation when it occurs, it would be more efficient (i.e., cost-effective) to abandon specific identifications and devote the saved effort to collecting more samples (Keough and Quinn 1991, Vanderklift et al. 1996). Ferraro and Cole (1995) summarize justifications for using only the level of taxonomy sufficient to detect the pollution effect of interest: (1) taxonomy costs would be minimized without

loss of precision or statistical rigour; (2) consistency between studies would be improved; (3) data quality would be improved because higher taxonomic levels tend to be easier to fix and less subject to revision; and (4) field studies would be completed faster. These conclusions are based on marine studies, in which strictly statistical methods of pollution assessment predominate. The question then becomes, what level of information is sufficient in freshwater biomonitoring?

Research and experience summarized earlier strongly suggest that higher levels of taxonomy are sufficient to distinguish marked environmental degradation, especially over large areas or differing habitats, but that subtle and small-scale disturbance is better detected by generic or specific identifications. Lower taxonomic levels also contain more ecological information that can be used to interpret the nature of the stress on the benthic community. A working assumption in this review is that biomonitoring for mine sites should be as sensitive as reasonably possible. For benthic macro-invertebrates, the level of reasonably complete taxonomy is relatively easy to define because there are clear break-points beyond which the effort required for further taxonomic penetration increases sharply. Hence, it seems sensible to apply the "lowest practical level" criterion to taxonomy for biomonitoring. Specific identifications might still be warranted in follow-up studies or surveys intended to examine a special problem more closely.

4.2.4 Mixed Taxonomy

One consequence of the common practice of identifying organisms to the lowest practical taxonomic rank is that the "species list" from any site contains a mixture of taxa, some as low as species, others at the family, order or class level. Example lists are presented in the Appendix. In a mixed taxonomy list, there may be an imbalance in the relative contribution of different taxa to distinguishing sites. For instance, a group of five congeneric species will weight the analysis more at the species level than at the genus level because of the influence of five species compared with one. This effect will be felt particularly in multivariate analyses. On the other hand, Waterhouse and Farrell (1985) submit that the greater weight of lower taxa in such a list is legitimate because species contribute more information than do larger groups.

Closely related to the issue of mixed taxonomy is the question of unidentified organisms. Most benthic samples contain some organisms, usually early instars of immature insects, that cannot be identified fully because of their small size and incomplete development. Others may be damaged, missing a key appendage or body part, or improperly preserved. At least four options are available for dealing with these organisms:

(1) delete them entirely from the sample;

(2) lump them all together in an "Other" category;

(3) apportion them among the identified organisms according to the ratio of abundances; or

(4) place them in the lowest taxon to which confident identification is possible.

The first two options are unsatisfactory, although sometimes used, because in the first option a bias is introduced into the sample and in the second the organisms contribute no useful information (beyond increasing the total density estimate), and complicate further analyses. The third option assumes that the ratio of taxa among unidentified organisms is the same as in the identified ones, which may or may not be true. The fourth option leads to an even greater mixing of taxonomic levels: organisms within the same list might be identified as *Baetis*, Baetidae, or even just Ephemeroptera. Nevertheless, taxonomists generally follow this option because it injects the least bias into the taxa list (e.g., Pettigrove 1990).

Given a taxa list set up following option 4, the statistician must then decide whether to delete the incompletely identified organisms (option 1 again), apportion the animals according to option 3 for analysis, or to raise all the identified animals to the level of the lowest fully identified taxon. For example, *Baetis*, *Cloeon* and *Pseudocloeon* might be all lumped together and analyzed as Baetidae (e.g., Ormerod 1987, Kerans et al. 1992). Which option is preferred for analysis depends on the distribution of the organisms among taxa. If identified specimens in the three species above constitute only 10% of the total, it makes sense to lump them together with the other 90% in the unidentified Baetidae. Conversely, if 90% of the identified organisms fall into the three genera, it is neater to apportion the remainder among them, and any resulting error will not seriously bias the analysis.

Unidentifiable organisms are a recurrent problem in any benthic invertebrate study, yet there is remarkably little guidance available on the subject. Only Fisheries and Oceans and Environment Canada (1995) have dealt with the issue, for the Environmental Effects Monitoring Program, and they provide no solid advice. This is an issue in need of more consideration.

4.2.5 Reference Collections

One final matter pertaining to taxonomy warrants mention. Many researchers have emphasized the importance of maintaining reference collections of invertebrates and of depositing voucher specimens with museums or other depositories (Resh and Unzicker 1975, Pettigrove 1990, Resh and McElravy 1993, Environment Canada 1993, Norris and Norris 1995). Reference collections allow future verifications of taxonomy, and facilitate

long-term comparisons of studies done by different workers at different times. Voucher specimens are especially important if species are differentiated (as species A, species B, etc.) but not named (Cranston 1990).

Many biologists who work on benthic invertebrates maintain a reference collection for their own benefit, but there is no organized effort to maintain standard reference collections for sites that are subject to routine biomonitoring. Efforts to establish reference collections for monitoring at mine sites should be encouraged. Such collections should be maintained by the mine or an independent third body and made available to researchers and consultants each time a benthic survey is carried out. Fresh specimens (voucher specimens) of each taxon should be added to the collection after each survey. If workers could also be persuaded to deposit specimens with museums, this might help foster closer links between taxonomic specialists and applied scientists.

4.3 Rare Species

Benthic invertebrate communities, like most animal communities, are composed of widely uneven numbers of component species. In a typical, healthy river, a few species are represented by many individuals, many species are represented by only a few individuals, and some species are intermediate. A plot of abundance class against numbers of individuals in each class usually resembles a log-normal distribution (Johnson et al. 1993). Figure 7 presents an idealized example of a log-normal species abundance distribution, while Figure 8 presents a couple of real examples. The shape of the species abundance curve may vary from site to site, and will also be affected by the level of taxonomic resolution (Figure 8), but in most streams and rivers the majority of species taken in any given sample are rare, collectively contributing <2% of the total number of individuals in the sample. Furthermore, because species distributions are patchy, replicate samples will not include exactly the same set of rare species. Consequently it is not practical to sample the benthic community exhaustively, nor is it possible to state definitively when all rare species have been sampled.

It must be stressed that in the present context, *rare* is defined in a strictly statistical sense. The term does not refer to species that are endangered, restricted in distribution, or otherwise of special conservation interest. Rare species (*undersampled* might be a more descriptive term) are those that are present in very low numbers in any sample, and for which population densities cannot be accurately estimated with a reasonable sampling effort. The species list for Blue Springs Creek in the Appendix, for example, contains a number of species with density estimates of 11 or 22 per square metre, meaning that only one or two individuals were trapped in a square-foot Surber sample. When numbers in a sample are that low, it is nearly impossible to detect a change in density from one site to another, especially given the extremely high variance relative to the mean. Even presence/absence data are of little value for species at the tail of the species distribution, because they may be missing in any given sample by chance alone.

Rare species may be an important component of the benthos community in terms of their interactions with other species and effects on ecosystem dynamics. Nevertheless, a strong case can be made for deleting rare species entirely from the species list in biomonitoring studies. The pragmatic and biological arguments supporting this suggestion are several:

(1) By definition, rare species will be those for which density estimates are least reliable, and hence for which differences between sites will be most difficult to detect. To achieve estimates of species population densities for rare species comparable for those of common species would require sampling intensity far beyond the reasonable limits of any biomonitoring program. Moreover, because the species distribution has a long tail of increasingly rare species, more intense sampling would add new species, even more rare, to the list, for which population estimates would still be imprecise. Hence, any sampling program is selective against some subset of rare species; deleting species below a predetermined threshold would make the selectivity explicit and fixed, instead of relying on chance.

(2) The effect of these species on results of statistical analyses is almost invariably small, yet they complicate or preclude the application of many methods because of low or zero counts in some replicates. The difficulty of detecting differences among sites using univariate comparisons of rare species has already been mentioned. Given the imprecision of density estimates for rare species, only a very large increase in population density can be detected, while a decrease, or even disappearance, cannot be confirmed.

In the calculation of similarity indices, and the ordination techniques that are built on them, rare species are literally more trouble than they are worth. As a consequence of low numbers, poor precision and random jumps in density from one sample to the next, rare species generally make only a meagre contribution to

discriminating sites or defining gradients (e.g., Pontasch et al. 1989). Yet their inclusion hampers the analysis because of the much larger matrices required and because zero entries render many computational methods cumbersome or unworkable. Correspondence analysis is very sensitive to the contribution of rare species, and the ordination axes may be affected by patterns in the presence or absence of rare taxa (Dolédec and Chessel 1991). In his standard text on ordination, Gauch (1982) recommends that rare species be deleted because they usually contribute nothing or behave as outliers. Multivariate analysis of variance (MANOVA) requires that the number of samples be larger than the number of species, and cannot even be considered for a species list with rare species included (Norris and Georges 1993).

(3) The abundant species contain most of the useful information in the sample, and with the exception of predators, abundant species more accurately reflect ecological conditions at the site: a rare species may be naturally rare, or may be living in conditions generally unsuitable for that species. Hence, knowledge of the ecological requirements of a rare species cannot be used with confidence to make inferences about conditions at the site where it was found. Note that this argument applies only to species that are taken in low numbers at every site. A species may be common at one site, but be statistically rare at an impaired site in response to toxins or disturbance.

The exception to this biological argument are big predators, among which the large-bodied stoneflies in the family Perlidae are a conspicuous example. Stoneflies tend to be present in low numbers even at healthy sites, and because they are predators, roaming about to find prey, they tend to be distributed more evenly than other insects (Morin 1985). Hence, a few predatory stoneflies per sample is the rule in productive streams even in the absence of pollution or disturbance. However, the practical arguments still apply. Though we know that these animals are sensitive to many kinds of pollution, it is difficult to draw conclusions from the presence or absence of perlid stoneflies at one site compared with another when the absolute numbers are low to begin with. Further, if stoneflies are interacting with other species at the site, then their contribution should be reflected in the abundances of other (prey) species (DFO & Environment Canada 1995).

The conclusion is that deletion of rare species greatly simplifies analysis without significant loss of information, and should be considered as a standard practice in benthic invertebrate biomonitoring. Until recently, most methods guides did not explicitly mention trimming species lists (Anderson 1990, Klemm et al. 1990). Nevertheless, among researchers, deletion of rare species is already routinely practised as a first step in data analysis (e.g., Rooke and Mackie 1982, Culp and Davies 1980, Pontasch and Brusven 1988, Pontasch et al. 1989 Whitehurst and Lindsey 1990). Making deletion of rare species a part of standard procedures would help biologists overcome the feeling that paring their species list is a sort of unsanctioned, clandestine activity.

Unfortunately, there has not yet emerged a uniform definition of rarity; researchers have used a variety of arbitrary cut-off points based on absolute density in single samples or across all samples, or based on the percentage composition of all samples pooled. Most commonly, species are deleted that constitute less than some arbitrary percentage (5%, 2% or 1%) of the total number of individuals in a sample, so as to truncate the species abundance curve (Figure 8). The guidance document for environmental effects monitoring at pulp mills recommends deleting taxa that both constitute <5% of total numbers and are found in only one sample (DFO & Environment Canada 1995); in this scheme rare species present at more than one site would remain. Others have taken the reverse approach, and retained the most common species until a specified proportion of the total individuals, usually 95%, were included (Environment Canada 1993, McCall and Soster 1996).

A uniform criterion for deleting species would be useful. An analysis of extant species lists would suffice to find the most efficient protocol. If common usage can be trusted, deletion of all species that compose <1% of total numbers from all sites combined appears to be a conservative rule that is gaining acceptance. However, the decision to delete would be better based on mean density in all replicates at each site, to avoid exaggerating variance, and more importantly, to avoid accidentally deleting species that are abundant at one site but rare at others.

Deleting rare species immensely simplifies analysis and presentation of benthic invertebrate data. Depending on the site and the criterion used, half to three quarters of the species in a full list might be considered rare (Figure 8A), Boulton 1985). Rare species probably demand a disproportionate amount of time for identification, because they represent many taxa and are likely to be less familiar than species that occur abundantly in every sample. If certain rare species are not going to contribute to the analysis, cost efficiency might be further improved by not taking the time to identify them. Instead, the taxonomist could separate and enumerate all taxa, but only identify the common ones, moving downward through the abundance classes until some predefined threshold (e.g., 95% of total numbers) had been passed. Whether this modification would lead to significant time savings in practice is questionable; individual species or taxa would still need to be separated from others, and in many groups the effort to do so, using taxonomic keys, would be equivalent to identifying each one. Still, the possibility is worthy of further consideration.

5. Conclusions and Recommendations

5.1 General Conclusions

Biomonitoring based on benthic macro-invertebrates in watercourses near Canadian mine sites can be improved by optimizing study design, field methods and laboratory methods to fit the specific, narrow objective of measuring changes in invertebrate community structure between sites, rather than attempting to emulate sampling procedures that have been developed for scientific studies or for inventories of the benthic community. The foregoing analysis of macro-invertebrate sampling methods based on that assumption leads to three unifying conclusions.

First and foremost, the problem of statistical power and minimum effect size must be resolved. The power of a sampling program is its capacity to detect environmental degradation, which in hypothesis testing is equivalent to avoiding a Type II error. Power is determined by the significance level used in the statistical test (α), the variability of the data, the number of replicates and the magnitude of the difference in question (Fairweather 1991). A monitoring program with low power will fail to detect impairment of the benthic invertebrate community at downstream stations relative to controls unless the difference is very large. In our preoccupation with designing programs that are robust (avoidance of Type I error), power is often forgotten.

In Section 2.2, it was mentioned that one drawback of sequential decision plans was that they required a minimum effect size, defined beforehand, that the plan would be designed to detect. The same limitation applies to the sampling optimization procedure of Ferraro et al. (1989, 1994) (Section 3.1.6). Fixed definitions of what constitutes a "significant effect" of an effluent have proved to be controversial, because some workers feel that such definitions arbitrarily decide that certain small effects are insignificant, when in fact their biological significance may be very real. Conventional biomonitoring studies have adopted a more exploratory approach in which any effect that was discernable by the study was considered to be biologically meaningful.

But in fact, beforehand decisions on what magnitude of effect will be detectable are made implicitly in every study by the choice of sample replication, sample size and significance level, which collectively determine the power of the study and the magnitude of the effect that can be detected. For years, statisticians have been urging biologists to explicitly consider power when designing monitoring studies (see Peterman 1990, Fairweather 1991 and references therein). If this advice were heeded, it would be apparent that sequential decision plans and Ferraro's sampling optimization procedure require no more assumptions about minimum effect sizes than any other monitoring program.

The second conclusion is that sampling bias is both inevitable and tolerable in any benthic invertebrate monitoring program. Bias is inevitable because organisms that are rare, small or cryptic, or cling tightly to the substratum or burrow deep within it, will always be missed. The undersampling of small species especially may be severe. But bias is still tolerable in a biomonitoring program because it is the differences among sites that matter. Bias should not hinder a sensitive biomonitoring program if (1) the bias is the same or nearly so at all sites and (2) enough of the indicative organisms are captured at each site. Smaller organisms missed in sampling probably do not contribute much to distinguishing sites; maintaining equal bias is the impetus behind careful selection of sampling sites and compensation for habitat factors discussed in Section 2.1.3.

The third conclusion is that optimization of sampling procedures would require a set of integrated, co-ordinated changes if the full value of these ideas is to be realized. If sample size is reduced, replication must be increased to ensure no loss of precision. Smaller samples take less time to sort and identify and should enable less reliance on subsampling, but the best laboratory facilitation methods might be different than for larger samples. The logistics of subcontracting to taxonomic specialists, who traditionally charge by the sample, must also be adjusted to allow for smaller size and the possibility that rare species do not need identification. All these changes require closer collaboration among field workers, taxonomists, statisticians and ecologists.

5.2 Recommendations

Study Design

(1) Biomonitoring studies at mine sites should incorporate two or more control sites wherever possible. Differences between benthic invertebrate communities at the control sites can be used to define the magnitude of natural variation, and help decide what size of change at downstream sites ought to be considered indicative of significant impairment. In most situations the return in information from more than two control sites probably does not justify the additional effort.

(2) Where an upstream control site is not possible because the mine discharges to a headwater stream or a lake or reservoir outlet, two alternative study designs are possible: (a) establish a site or sites on a comparable nearby stream; (b) establish a baseline of information from many streams in the region (reference sites) for comparison against the study stream. The problem of mine sites at headwaters is incompletely resolved and deserving of further examination.

(3) At least the following habitat variables should be measured at every site:

- water depth
- water velocity
- substratum particle size
- standing crop of algae or detritus (flowing waters)
- total organic carbon (standing waters)

The best way to incorporate these data into the analysis (regressions, ordinations, analysis of covariance) is an important unresolved issue.

(4) The best approach presently available for assessing the effects of multiple effluents and non-point sources is to sample invertebrates above and below each outfall. Toxicity tests on effluents, plume delineation studies, and tracer chemicals can help unravel the contribution of different sources, but increase the complexity and expense of the study. It is not always possible to determine the presence, nature and extent of all the impairment and recovery zones in a river receiving multiple effluents. This is a major and widespread problem that should be addressed soon.

Sequential Decision Plans

(5) The utility and practicality of sequential decision plans for biomonitoring at mine sites should be examined and tested. Decision plans can only be used if a minimum effect size is agreed upon, and the approximate sampling distribution of the variable of interest is known.

Rapid Assessment Procedures

(6) Rapid assessment procedures are too insensitive to be useful in most routine mining monitoring, but they may occasionally be useful for confirmation of severe impairment. Many of the metrics used in rapid assessment approaches are equally applicable to conventional statistical analysis. Therefore, research on rapid assessment procedures may produce useful ideas for conventional parametric biomonitoring.

Sample Size and Replication

(7) Cost efficiency of benthic invertebrate monitoring programs would be dramatically improved by using much smaller samplers and increasing the number of replicates at each site. For stream sampling, devices such

as the T-sampler, which sample an area of 100 cm^2, should be strongly preferred over conventional devices such as the Surber sampler, which sample an area 10 times larger. The effort saved from collecting smaller samples should be devoted to increasing the number of replicates from the present level of five or less to 10 or more per station. A comparison of the efficacy of small with large samplers at mine sites is recommended.

Mesh Size

(8) Small animals, early instars of larval insects, and especially chironomidae are severely undersampled by mesh sizes of 500 μm or larger. To sample these organisms accurately would require extremely fine meshes that are not practical or cost-effective for biomonitoring. A mesh size of 250 μm is the best compromise between efficiency and reasonably complete retention of most macro-invertebrates, and is recommended for biomonitoring at mines. Ensuring that different investigators use the same mesh size at a given site is at least as important as the actual mesh size used.

Sampler Bias

(9) Sampler bias is unavoidable but does not impede detection of site differences if the bias is equal among sites and most larger species are sampled adequately. Differences in sampler bias among sites can be minimized by careful site selection, measurement of physical habitat variables, and collection of samples at all sites by one or two trained individuals.

Sample Sorting

(10) All of the various methods for facilitating sorting work to some degree, and all have limitations. Facilitation methods are valuable time savers and sharply improve the cost-efficiency of sorting benthic samples, if minimum standards of specimen recovery can be met. Workers should be encouraged to use extant methods routinely and to test and apply new ideas.

(11) Subsampling increases the imprecision of density estimates and should be used only where necessary. If samples were smaller subsampling would be needed much less often. Most workers are aware of the need to avoid bias and take as large a subsample as possible. Some loss of information about rarer species is inevitable when subsampling is employed.

Taxonomic Resolution

(12) Identifications of specimens to the lowest practical level, which equates with genus for most insects and the lowest level possible without special procedures (dissection, microscopy) or reliance on specialists for all other groups, is sufficient for biomonitoring in the mining industry. The minimum level of taxonomic resolution for biomonitoring should be specified, to encourage uniformity of practice. More complete taxonomy, even to species for some insects, may be warranted in follow-up studies or surveys intended to examine a special problem more closely if the added information justifies the higher cost.

(13) Mixed taxonomy and unidentifiable organisms are a ubiquitous problem in benthos samples, and solid guidance on how best to deal with these taxa is sorely needed. A simple desktop study, using extant species lists, to explore the effects of mixed taxonomy and different methods for dealing with it on the precision and accuracy of monitoring studies, is recommended.

(14) Reference collections, preferably maintained by an independent body, can help taxonomists with identifications and ensure uniform and comparable taxonomy between workers and over time. A reference collection of benthic invertebrates should be maintained for every mine site and should be made available to consultants or researchers when each biomonitoring study is undertaken. Voucher specimens should be deposited in the reference collection after each survey. Closer cooperation between museums, taxonomic experts and workers carrying out biomonitoring studies should be actively encouraged.

Rare Species

(15) Deletion of statistically rare species, those for which the estimate of mean density is too imprecise to be useful, greatly simplifies analysis without significant loss of information, and should be considered as a standard practice in benthic invertebrate biomonitoring. A uniform criterion for deleting species would greatly simplify this procedure. A desk-top analysis of extant species lists from a wide variety of lotic and lentic sites should be undertaken to find the most efficient and widely applicable protocol for deciding which species to delete. The possibility of further time savings by omitting identifications of rare species also warrants investigating.

6. References

Agard, J.B.R., Gobin, J. and Warwick, R.M. 1993. Analysis of marine macrobenthic community structure in relation to pollution, natural oil seepage and seasonal disturbance in a tropical environment (Trinidad, West Indies). Mar. Ecol. Prog. Series. 92: 233-243.

Allan, R.K. 1984. A new classification of the sub-family Ephemerellinae and the description of a new genus. Pan-Pac. Entomol. 60: 245-247.

Anderson, A.M. 1990. Selected methods for the monitoring of benthic invertebrates in Alberta rivers. Environmental Quality Monitoring Branch, Environmental Assessment Division, Alberta Environment, Edmonton, Alberta. 41 p.

Anderson, R.O. 1959. A modified flotation technique for sorting bottom fauna samples. Limnol. Oceanogr. 4: 223-225.

APHA (American Public Health Association). 1992. Standard methods for the examination of water and wastewater. 18th Edition. Washington, D.C.

Armitage, P.D. 1977. Invertebrate drift in the regulated River Tees and an unregulated tributary Maize Beck, below Cow Green Dam. Freshwat. Biol. 7: 167-183.

Armitage, P.D. 1978. Catches of invertebrate drift by pump and net. Hydrobiologia 60: 229-233.

Armitage, P.D., and Gunn, R.J.M. 1996. Differential response of benthos to natural and anthropogenic disturbances in 3 lowland streams. Int. Revue ges. Hydrobiol. 81: 161-181.

Armitage, P.D., Gunn, R.J.M., Furse, M.T., Wright, J.F. and Moss, D. 1987. The use of prediction to assess macroinvertebrate response to river regulation. Hydrobiologia 144: 25-32.

Bachalet, G. 1990. The choice of sieving mesh size in the quantitative assessment of marine macrobenthos: A necessary compromise between aims and constraints. Mar. Envir. Res. 30: 21-35.

Barbour, M.T., Gerritsen, J., Griffith, G.E., Frydenborg, R., McCarron, E., White, J.S. and Bastian, M.L. 1996. A framework for biological criteria for Florida streams using benthic macroinvertebrates. J. N. Amer. Benthol. Soc. 15(2): 185-211.

Barbour, M.T., Plafkin, J.L., Bradley, B.P., Graves, C.G. and Wisseman, R.W. 1992. Evaluation of EPA's rapid bioassessment benthic metrics: Metric redundancy and variability among reference stream sites. Envir. Toxicol. Chem. 11: 437-499.

Barmuta, L.A. 1984. A method of separating benthic arthropods from detritus. Hydrobiologia 112: 105-108.

Boulton, A.J. 1985. A sampling device that quantitatively collects benthos in flowing or standing waters. Hydrobiologia 127: 31-40.

Bournaud, M. et al. 1996. Macroinvertebrate community structure and environmental characteristics along a large river: Congruity of patterns for identification to species or family. J. N. Amer. Benthol. Soc. 15(2): 232-253.

Brinkman, M.A. and Duffy, W.G. 1996. Evaluation of four wetland aquatic invertebrate samplers and four sample sorting methods. J. Freshwat. Ecol. 11: 193-200.

Brooks, S. 1994. An efficient and quantitative aquatic benthos sampler for use in diverse habitats with variable flow regimes. Hydrobiologia 281-123-128.

Brown, A.V., Schram, M.D. and Brussock, P.P. 1987. A vacuum benthos sampler suitable for diverse habitats. Hydrobiologia 153: 241-248.

Bunn, S.E. 1995. Biological monitoring of water quality in Australia: Workshop summary and future directions. Aust. J. Ecol. 20: 220-227.

Cairns, J. Jr. and Dickson, K.L. 1971. A simple method for the biological assessment of the effects of waste discharges on aquatic bottom dwelling organisms. J. Water. Pollut. Control Fed. 43: 755-772.

Cash, K.J. 1995. Assessing and monitoring aquatic ecosystem health: Approaches using individual population and community/ecosystem measurements. Northern River Basins Study Project Report No. 45, Edmonton, Alberta. 70 p.

Camacho, F. and Vascotto, G.L. 1991. Framework for enhancing the statistical design of aquatic environmental studies. Environ. Monitor. Assess. 17: 303-314.

Chessman, B.C. 1995. Rapid assessment of rivers using macroinvertebrates: A procedure based on habitat-specific sampling, family-level identification and a biotic index. Aust. J. Ecol. 20: 122-129.

Chutter, F.M. 1975. Variation in the day-time drift of a Natal river. Verh. int. Ver. Limnol. 19: 1728-1735.

Ciborowski, J.J.H. 1991. Estimating processing time of stream benthic samples. Hydrobiologia 222: 101-107.

Clifford, H.F. 1972a. A year's study of the drifting organisms in a brown-water stream of Alberta, Canada. Can. J. Zool. 50: 975-983.

Clifford, H.F. 1972b. Drift of invertebrates in an intermittant stream draining marshy terrain of west-central Alberta. Can. J. Zool. 50: 985-991.

Clifford, H.F. and Casey, R.J. 1992. Differences between operators in collecting quantitative samples of stream macroinvertebrates. J. Freshwat. Ecol. 7: 271-276.

Corkum, L.D. 1990. Intrabiome distributional patterns of lotic macroinvertebrate assemblages. Can. J. Fish. Aquat. Sci. 47: 2147-2157.

Corkum, L.D. and Ciborowski, J.J.H. 1998. Use of alternative classifications in studying broad-scale distributional patterns of lotic invertebrates. J. N. Amer. Benthol. Soc. 7: 167-179.

Cranston, P.S. 1990. Biomonitoring and invertebrate taxonomy. Environ. Monitor. Assess. 14: 265-273.

Cromar, G.L. and Williams, D.D. 1991. Centrifugal flotation as an aid to separating invertebrates from detritus. Hydrobiologia 209: 67-70.

Courtemanch, D.L. 1996. Commentary on the subsampling procedures used for rapid bioassessments. J. N. Amer. Benthol. Soc. 15: 381-385.

Culp, J.M. and Davies, R.W. 1980. Reciprocal averaging and polar ordination as techniques for analysing lotic macroinvertebrate communities. Can. J. Fish. Aquat. Sci. 37: 1358-1364.

Delong, M.D., Thorp, J.W. and Haag, K.H. 1993. A new device for sampling macroinvertebrates from woody debris (snags) in nearshore areas of aquatic systems. Amer. Midl. Nat. 130: 413-417.

DFO (Department of Fisheries and Oceans) and Environment Canada. 1993. Technical guidance document for aquatic environmental effects monitoring related to federal *Fisheries Act* requirements. Version 1.0. Environmental Effects Monitoring Program, Environmental Conservation Service, Environment Canada, Ottawa, Ontario. 128 p.

DFO (Department of Fisheries and Oceans) and Environment Canada. 1995. Further guidance for the invertebrate community survey for aquatic environmental effects monitoring related to the federal *Fisheries Act* requirements. Environmental Effects Monitoring Program, Environmental Conservation Service, Environment Canada, Ottawa, Ontario. Report EEM 2. 206 p.

Doeg, T. and Lake, P.S. 1981. A technique for assessing the composition and density of the macroinvertebrate fauna of large stones in streams. Hydrobiologia 80: 3-6.

Dolédec, S. and Chessel, D. 1991. Recent developments in linear ordination methods for environmental sciences. Adv. Ecol. 1: 133-155.

Downing, J.A. 1979. Aggregation, transformation, and the design of benthos sampling programs. J. Fish. Res. Bd. Can. 36: 1454-1463.

Downing, J.A. 1980. Corrections to recent publications. Can. J. Fish. Aquat. Sci. 37: 1333.

Downing, J.A. 1984. Sampling the benthos of standing waters. In: Downing, J.A. and Rigler, F.H. (editors). A manual on methods for the assessment of secondary productivity in fresh waters. Second ed. IBP Handbook 17, Blackwell Scientific Pubs., Oxford, England. p. 87-130. (Cited in Resh and McElravy 1993)

Downing, J.A. and Anderson, M.A. 1985. Estimating the standing biomass of aquatic macrophytes. Can. J. Fish. Aquat. Sci. 42: 1570-1579.

Downing, J.A. and Cyr, H. 1985. Quantitative estimation of epiphytic invertebrate populations. Can. J. Fish. Aquat. Sci. 42: 1570-1579.

Downing, J.A. 1989. Precision of the mean and the design of benthos sampling programmes: Caution revised. Mar. Biol. 103: 231-234.

Elliott, J.M. 1977. Some methods for the statistical analysis of samples of benthic invertebrates. Second ed. Pub. No. 25, Freshwater Biological Association (UK). 160 p.

Elliott, J.M. and Tullett, P.A. 1978. A bibliography of samplers for benthic invertebrates. Occasional Paper No. 4, Freshwater Biological Association (UK). 61 p.

Elliott, J.M. and Tullett, P.A. 1983. A supplement to a bibliography of samplers for benthic macroinvertebrates. Occasional Paper No. 20, Freshwater Biological Association (UK). 27 p.

Erman, D.C. 1981. Stream macroinvertebrate baseline surveys: A comparative analysis from the oil-shale regions of Colorado, U.S.A. Environ. Manage. 5: 531-536.

Environment Canada. 1993. Guidelines for monitoring benthos in freshwater environments. Environment Canada, North Vancouver, B.C. 81 p.

Fairweather, P.G. 1991. Statistical power and design requirements for environmental monitoring. Aust. J. Mar. Freshwat. Res. 42: 555-567.

Faith, D.P., Dostine, P.L. and Humphrey, C.L. 1995. Detection of mining impacts on aquatic macroinvertebrate communities: Results of a disturbance experiment and the design of a multivariate BACIP monitoring programme at Coronation Hill, Northern Territory. Aust. J. Ecol. 20: 167-180.

Faith, D.P., Humphrey, C.L. and Dostine, P.L. 1991. Statistical power and BACI designs in biological monitoring: Comparative evaluation of measures of community dissimilarity based on benthic macroinvertebrate communities in Rockhole Mine Creek, Northern Territory, Australia. Aust. J. Mar. Freshwat. Res. 42: 589-602.

Fernet, D.A. and Walder, G.L. 1986. The development of habitat suitability criteria for use with the instream flow incremental methodology. A manual for the Province of Alberta. Habitat Branch, Fish and Wildlife Division, Edmonton, Alberta. 141 p.

Ferraro, S.P. and Cole, F.A. 1995. Taxonomic level sufficient for assessing pollution impacts on the Southern California Bight macrobenthos -- Revisited. Environ. Toxicol. Chem. 14: 1031-1040.

Ferraro, S.P., Cole, F.A., DeBen, W.A. and Swartz, R.C. 1989. Power-cost efficiency of eight macrobenthic sampling schemes in Puget Sound, Washington, U.S.A. Can. J. Fish. Aquat. Sci. 46: 2157-2165.

Ferraro, S.P., Swartz, R.C., Cole, F.A. and Deben, W.A. 1994. Optimum macrobenthic sampling protocol for detecting pollution impacts in the Southern California Bight. Environ. Monitor. Assess. 29: 127-153.

Furse, M.T., Moss, D., Wright, J.F. and Armitage, P.D. 1984. The influence of seasonal and taxonomic factors on the ordination and classification of running-water sites in Great Britain and on the prediction of their macro-invertebrate communities. Freshwat. Biol. 14: 257-280.

Gauch, H.G. Jr. 1982. Multivariate analysis in community ecology. Cambridge University Press, Cambridge, U.K. 298 p.

Golder Associates Ltd. 1993. Joint Industry-Municipal North Saskatchewan River Study. Unpublished report for Capital Region Sewage Commission (Edmonton), City of Edmonton, and nine Edmonton industries. Golder Associates, Calgary, Alberta. 122 p.

Gray, J.S., Aschan, M., Carr, M.R., Clarke, C.R., Green, R.H., Pearson, T.H., Rosenberg, R. and Warwick, R.M. 1988. Analysis of community attributes of the benthic macrofauna of Frierfjord/ Langesundfjord and in a mesocosm experiment. Mar. Ecol. Prog. Ser. 67: 251-262.

Gray, J.S., Clarke, K.R., Warwick, R.M. and Hobbs, G. 1990. Detection and initial effects of pollution on marine benthos: An example from the Ekofisk and Eldfisk oilfields, North Sea. Mar. Ecol. Prog. Ser. 66: 285-299.

Green, R.H. 1979. Sampling design and statistical methods for environmental biologists. John Wiley & Sons, New York, NY. 257 p.

Growns, I.O. 1990. Efficiency estimates of a stream benthos suction sampler. Aust. J. Mar. Freshwat. Res. 41: 621-626.

Growns, J.E., Chessman, B.C., McEvoy, P.K. and Wright, I.A. 1995. Rapid assessment of rivers using macroinvertebrates: Case studies in the Nepean River and Blue Mountains, NSW. Aust. J. ecol. 20: 130-141.

Hall, R.J. and Ide, F.P. 1987. Evidence of acidification effects on stream insect communities in central Ontario between 1937 and 1985. Can. J. Fish. Aquat. Sci. 44: 1652-1657.

Harper, P.P. and Cloutier, L. 1986. Spatial structure of the insect community of a small dimictic lake in the Laurentians (Québec). Int. Rev. ges. Hydrobiol. 71: 655-685.

Hauer, F.R. and Lamberti, G.A. (editors). 1996. Methods in stream ecology. Academic Press Inc., San Diego California. 674 p.

Heip, C., Warwick, R.M., Carr, M.R., Herman, P.M.J., Huys, R., Smol, N. and Van Holsbeke, K. 1988. Analysis of community attributes of the benthic meiofauna of Frierfjord/Langesundfjord. Mar. Ecol. Prog. Ser. 46: 171-180.

Hellawell, J.M. 1977. Change in natural and managed ecosystems: Detection, measurement and assessment. Proc. Roy. Soc. London (B) 197: 31-57.

Herricks, E.E. and Cairns, J. Jr. 1982. Biological monitoring. Part III -- Receiving system methodology based on community structure. Water Research 16: 141-153.

Hickley, P. 1975. An apparatus for subdividing benthos samples. Oikos 26: 92-96.

Hilsenhoff, W.L. 1987. An improved biotic index of organic stream pollution. Great Lakes Entomol. 20: 31-39.

Hilsenhoff, W.L. 1988. Rapid field assessment of organic pollution with a family-level biotic index. J. N. Amer. Benthol. Soc. 7: 65-68.

Hughes, B.D. 1978. The influence of factors other than pollution on the value of Shannon's diversity index for benthic macro-invertebrates in streams. Water Research 12: 359-364.

Humphrey, C.L., Faith, D.P. and Dostine, P.L. 1995. Baseline requirements for assessment of mining impact using biological monitoring. Aust. J. Ecol. 20: 150-166.

Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments. Ecol. Monogr. 54: 187-211.

Jackson, J.K. and Resh, V.H. 1988. Sequential decision plans in monitoring benthic macroinvertebrates: Cost savings, classification accuracy and development of plans. Can. J. Fish. Aquat. Sci. 45: 280-286.

Jackson, J.K. and Resh, V.H. 1989. Sequential decision plans, benthic macroinvertebrates, and biological monitoring programs. Envir. Manage. 13: 455-468.

Johnson, R.K., Widerholm, T. and Rosenberg, D.M. 1993. Freshwater biomonitoring using individual organisms, populations, and species assemblages of benthic macroinvertebrates. In: Rosenberg, D.M. and Resh, W.H. (editors). Freshwater biomonitoring and benthic macroinvertebrates. Chapman & Hall, New York, N.Y. p. 40-158.

Jónasson, P.M. 1958. The mesh factor in sieving techniques. Verh. internat. Ver. Limnol. 13: 860-866.

Kaesler, R.L. and Herricks, E.E. 1980. Discussion: "Hierarchial diversity of communities of aquatic insects and fishes," by Roger L. Kaesler and Edwin E. Herricks. Water Resour. Bull. 16: 366-367.

Keough, M.J. and Quinn, G.P. 1991. Causality and the choice of measurements for detecting human impacts in marine environments. Aust. J. Mar. Freshwat. Res. 42: 539-554.

Kerans, B.L., Karr, J.R., and Ahlstedt, S.A. 1992. Aquatic invertebrate assemblages - spatial and temporal differences among sampling protocols. J. N. Amer. Benthol. Soc. 11(4): 377-390.

Klemm, D.J., Lewis, P.A., Fulk, F. and Lazorchak, J.M. 1990. Macroinvertebrate field and laboratory methods for evaluating the biological integrity of surface waters. Office of Research and Development, United States Environmental Protection Agency, Washington, D.C. EPA/600/4-90/030. 256 p.

Kroger, R.L. 1972. Underestimation of standing crops by the Surber sampler. Limnol. Oceanogr. 17: 475-478.

Lackey, R.T. and May, B.E. 1971. Use of sugar flotation and dye to sort benthic samples. Trans. Amer. Fish. Soc. 100: 794-797.

Lenat, D.R. and Penrose, D.L. 1980. Discussion: "Hierarchial diversity of communities of aquatic insects and fishes," by Roger L. Kaesler and Edwin E. Herricks. Water Resour. Bull. 16: 361-362.

Long, B.G. and Wang, Y.G. 1994. Method for comparing the capture efficiency of benthic sampling devices. Mar. Biol. 121: 397-399.

Lund, J.W.G., Kipling, C. and Le Cren, E.S. 1958. The inverted microscope method of estimating algal numbers and the statistical basis of estimations by counting. Hydrobiologia 11: 143-170.

Mackie, G.L. and Bailey, R.C. 1981. An inexpensive stream-bottom sampler. J. Freshwat. Ecol. 1: 61-69.

Magdych, W.P. 1981. An efficient, inexpensive elutriator design for separating benthos from sediment samples. Hydrobiologia 85: 157-160.

Magdych, W.P. 1984. Salinity stress along a complex river continuum: Effects on mayfly (Ephemeroptera) distributions. Ecology 65: 1662-1672.

Marchant, R.M. 1990. Robustness of classifications and ordination techniques applied to macroinvertebrate communities from the La Trobe River, Victoria. Aust. J. Mar. Freshwat. Res. 41: 493-504.

Marchant. R., Barmuta, L.A. and Chessman, B.C. 1995. Influence of sample quantification and taxonomic resolution on the ordination of macroinvertebrate communities from running waters in Victoria, Australia. Mar. Freshwat. Res. 46: 501-506.

Mason, W.T. and Yevich, P.P. 1967. The use of phloxine B and rose bengal stains to facilitate sorting benthic samples. Trans. Amer. Microsc. Soc. 56: 221-223.

Mason, W.T., Lewis, P.A. and Weber, C.I. 1985. An evaluation of benthic macroinvertebrate biomass methodology. Part 2. Field assessment and data evaluation. Envir. Monitor. Assess. 5: 399-422.

Mccafferty, W.P. 1992. New larval descriptions and comparisons of North American *Chloroterpes* (Ephemeroptera, Leptophlebidae). Great Lakes Entomol. 25: 71-78.

Mccafferty, W.P. and Waltz, R.D. 1990. Revisionary synopsis of the Baetidae (Ephemeroptera) of North and Middle America. Trans. Amer. Entomol. Soc. 116: 769-800.

Merritt, R.W. and Cummins, K.W. (editors) 1984. An introduction to the aquatic insects of North America. Second edition. Kendall Hunt Pub. Co., Iowa. 722 p.

Merritt, R.W., Cummins, K.W. and Resh, V.H. 1984. Collecting, sampling and rearing methods for aquatic insects. In: Merritt, R.W. and Cummins, K.W. (editors) An introduction to the aquatic insects of North America. Second edition. Kendall Hunt Pub. Co., Iowa. p. 11-26.

Metcalfe, J.L. 1989. Biological water quality assessment of running waters based on macroinvertebrate communities: History and present status in Europe. Envir. Pollut. 60: 101-119.

Meyer, E. 1990. A simple subsampling device for macroinvertebrates with general remarks on the processing of stream benthos samples. Arch. Hydrobiol. 117: 309-318.

Millard, S.P., Yearsley, J.R. and Lettenmaier, D.P. 1985. Space-time correlation and its effects on methods for detecting aquatic ecological change. Can. J. Fish. Aquat. Sci. 42: 1391-1400. (Correction in Can. J. Fish. Aquat. Sci. 43: 1680)

Morin, A. 1985. Variability of density estimates and the optimization of sampling programs for stream benthos. Can. J. Fish. Aquat. Sci. 42: 1530-1534.

Moss, D., Furse, M.T., Wright, J.F. and Armitage, P.D. 1987. The prediction of the macro-invertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. Freshwat. Biol. 17: 41-52.

Mundie, J.H. 1971. Sampling benthos and substrate materials, down to 50 microns in size, in shallow streams. J. Fish. Res. Board Can. 28: 849-860.

Nalepa, T.F. and Robertson, A. 1981. Screen mesh size affects estimates of macro- and meio-benthos abundance and biomass in the Great Lakes. Can. J. Fish. Aquat. Sci. 38: 1027-1034.

Norris, R.H. and Georges, A. 1993. Analysis and interpretation of benthic macroinvertebrate surveys. In: Rosenberg, D.M. and Resh, V.H. (editors) 1993. Freshwater biomonitoring and benthic macroinvertebrates. Chapman & Hall, New York, NY. p. 234-286.

Norris, R.H. and Norris, K.R. 1995. The need for biological assessment of water quality: Australian perspective. Aust. J. Ecol. 20: 1-6.

Ormerod, S.J. 1987. The influences of habitat and seasonal sampling regimes on the ordination and classification of macroinvertebrate assemblages in the catchment of the River Wye, Wales. Hydrobiologia 150: 143-151.

Osborne, L.L., Davies, R.W. and Linton, K.J. 1980. Use of hierarchial diversity indices in lotic community analysis. J. Appl. Ecol. 17: 567-580.

Pescador, M.L. 1985. Systematics of the nearctic genus *Pseudiron* (Ephemeroptera: Heptageniidae: Pseudironinae). Florida Entomol. 68: 432-443.

Peterman, R.M. 1990. Statistical power analysis can improve fisheries research and management. Can. J. Fish. Aquat. Sci. 47: 2-15.

Pettigrove, V. 1990. The importance of site selection in monitoring the macroinvertebrate communities of the Yarra River, Victoria. Environ. Monitor. Assess. 14: 297-313.

Plafkin, J.L., Barbour, M.T., Porter, K.D., Gross, S.K. and Hughes, R.M. 1989. Rapid bioassessment protocols for use in streams and rivers. Benthic macroinvertebrates and fish. Office of Water Regulation and Standards, United States Environmental Protection Agency, Washington, D.C. EPA/440/4-89/001

Pontasch, K.W. and Brusven, M.A. 1988. Diversity and community comparison indices: Assessing macroinvertebrate recovery following a gasoline spill. Water Research 22: 619-626.

Pontasch, K.W., Smith, E.P. and Cairns, J. Jr. 1989. Diversity indices, community comparison indices and canonical discriminant analysis: Interpreting the results of multispecies toxicity tests. Water Research 23: 1229-1238.

Pringle, J.D. 1984. Efficiency estimates for various quadrat sizes used in benthic sampling. Can. J. Fish. Aquat. Sci. 41: 1485-1489.

Provonsha, A.V. 1990. A revision of the genus *Caenis* in North America. Trans. Amer. Entomol. Soc. 116: 801-820.

Reger, S.J., Brothersen, C.F., Osborn, T.G. and Helm, W.T. 1982. Rapid and effective processing of macroinvertebrate samples. J. Freshwat. Ecol. 1: 451-466.

Reish, D.J. 1959. A discussion of the importance of the screen size in washing quantitative marine bottom samples. Ecology 40: 307-309.

Resh, V.H. 1979. Sampling variability and life history features: basic considerations in the design of aquatic insect studies. J. Fish. Res. Board Can. 36: 290-311.

Resh, V.H. and Jackson, J.K. 1993. Rapid assessment approaches to biomonitoring using benthic macroinvertebrates. In: Rosenberg, D.M. and Resh, V.H. (editors) 1993. Freshwater biomonitoring and benthic macroinvertebrates. Chapman & Hall, New York, NY. p. 195-233.

Resh, V.H., Jackson, J.K. and McElravy, E.P. 1988. The use of long-term ecological data and sequential decision plans in monitoring the impact of geothermal energy development on benthic macroinvertebrates. Verh. Internat. Verein. Limnol. 23: 1142-1146.

Resh, V.H. and McElravy, E.P. 1993. Contemporary quantitative approaches to biomonitoring using benthic macroinvertebrates. In: Rosenberg, D.M. and Resh, V.H. (editors) 1993. Freshwater biomonitoring and benthic macroinvertebrates. Chapman & Hall, New York, NY. p. 159-194.

Resh, W.H., Norris, R.H. and Barbour, M.T. 1995. Design and implementation of rapid assessment approaches for water resource monitoring using benthic macroinvertebrates. Aust. J. Ecol. 20: 108-121.

Resh, V.H. and Price, D.G. 1984. Sequential sampling: A cost-effective approach for monitoring benthic macroinvertebrates in environmental impact assessments. Environ. Manage. 8: 75-80.

Resh, V.H., Rosenberg, D.M. and Feminella, J.W. 1985. The processing of benthic samples: Responses to the 1983 NABS questionnaire. Bull. North Amer. Benthol. Soc. 2: 5-11. (Cited in Resh and McElravy 1993)

Resh, V.H. and Unzicker, J.D. 1975. Water quality monitoring and aquatic organisms: Importance of species identification. J. Water Pollut. Control Fed. 47: 9-19.

Reynoldson, T.B., Bailey, R.C., Day, K.E. and Norris, R.H. 1995. Biological guidelines for fresh-water sediment based on Benthic Assessment of Sediment (the *Beast*) using a multivariate approach for predicting biological state. Aust. J. Ecol. 20: 198-219.

Riddle, M.J. 1989. Precision of the mean and the design of benthos sampling programmes: Caution advised. Mar. Biol. 103: 225-230.

Robertson, D.J. and Piwowar, K. 1985. Comparison of four samplers for evaluating macroinvertebrates of a sandy gulf coast plain stream. J. Freshwat. Ecol. 3: 223-231.

Rooke, J.B. and Mackie, G.L. 1982. An ecological analysis of lotic environments: I. Design and testing. J. Freshwat. Ecol. 1: 433-441.

Rosenberg, D.M. and Resh, V.H. (editors). 1993. Freshwater biomonitoring and benthic macroinvertebrates. Chapman & Hall, New York, NY. 488 p.

Rossillon, D. 1987. About the separation of benthos from stream samples. Arch. Hydrobiol. 110: 469-475.

Schlacher, T.A. and Wooldridge, T.H. 1996. How accurately can retention of benthic macrofauna by a particular mesh size be predicted from body size of organisms? Hydrobiologia 323: 149-154.

Scrimgeour, G.J., Culp, J.M. and Glozier, N.E. 1993. An improved technique for sampling lotic invertebrates. Hydrobiologia 254: 65-

Sebastien, R.J., Rosenberg, D.M. and Wiens, A.P. 1988. A method for subsampling unsorted benthic macroinvertebrates by weight. Hydrobiologia 157: 69-75.

Sell, D.W. and Evans, M.S. 1982. A statistical analysis of subsampling and an evaluation of the Folsom plankton splitter. Hydrobiologia 94: 223-230.

Sheldon, A.L. 1984. Cost and precision in a stream sampling program. Hydrobiologia 111: 147-152.

Shipley, F.S. 1987. Sampling effort and bias in soft-sediment benthic investigations using the Peterson dredge. Texas J. Science 39: 118-122.

Slack et al. 1991. Mesh-size effects on drift sample composition as determined with a triple-net sampler. Hydrobiologia 209: 215-226.

Smith, S.D.A. and Simpson, R.D. 1993. Effects of pollution on holdfast macrofauna of the kelp *Ecklonia radiata*: Discrimination at different taxonomic levels. Mar. Ecol. Prog. Ser. 46: 167-170.

Smith, E.P., Orvos, D.R. and Cairns, J. Jr. 1993. Impact assessment using the before-after-control-impact (BACI) model: Concerns and comments. Can. J. Fish. Aquat. Sci. 50: 627-637.

Stewart-Oaten, A., Murdoch, W.W. and Parker, K.R. 1986. Environmental impact assessment: "Pseudoreplication" in time? Ecology 67: 929-940.

Stewart-Oaten, A., Bence, J.R. and Osenberg, C.W. 1992. Assessing effects of unreplicated perturbations: No simple solutions. Ecology 73: 1396-1404.

Storey, A.W. and Pinder, L.C.V. 1985. Mesh-size and efficiency of sampling of larval Chironomidae. Hydrobiologia 124: 193-198.

Taylor, B.R. and Roff, J.C. 1986. Long-term effects of highway construction on the ecology of a southern Ontario stream. Environ. Pollut. A: 40: 317-344.

Underwood, A.J. 1991. Beyond BACI: Experimental designs for detecting human environmental impacts on temporal variations in natural populations. Aust. J. Mar. Freshwat. Res. 42: 569-587.

Underwood, A.J. 1992. Beyond BACI: The detection of environmental impacts on populations in the real, but variable world. J. Exp. Mar. Biol. Ecol. 161: 145-178.

Underwood, A.J. 1993. The mechanics of spatially replicated sampling programmes to detect environmental impacts in a variable world. Aust. J. Ecol. 18: 99-116.

Vanderklift, M.A., Ward, T.J. and Jacoby, C.A. 1996. Effect of reducing taxonomic resolution on ordinations to detect pollution-induced gradients in macrobenthic infaunal assemblages. Mar. Ecol. Prog. Ser. 136: 137-145.

Vannote, R.L., Minshall, G.W., Cummins, K.W., Sedell, J.R. and Cushing, C.E. 1980. The river continuum concept. Can. J. Fish. Aquat. Sci. 37: 130-137.

Veijola, H., Meriläinen, J.J. and Marttila, V. 1996. Sample size in the monitoring of benthic macrofauna in the profundal of lakes: Evaluation of the precision of estimates. Hydrobiologia 322: 301-315.

Vézina, A.F. 1988. Sampling variance and the design of quantitative surveys of the marine benthos. Mar. Biol. 97: 151-155.

Vinson, M.R. and Hawkins, C.P. 1996. Effects of sampling area and subsampling procedure on comparisons of taxa richness among streams. J. North Amer. Benthol. Soc. 15: 392-398.

Voshell, J.R. Jr., Hiner, S.W. and Layton, R.J. 1992. Evaluation of a benthic macroinvertebrate sampler for rock outcrops in rivers. J. Freshwat. Ecol. 7: 1-6.

Voshell, J.R. Jr., Layton, R.J. and Hiner, S.W. 1989. Field techniques for determining the effects of toxic substances on benthic macroinvertebrates in rocky-bottomed streams. In: Cowgill, U. and Williams, L.R. (editors). Aquatic toxicology and hazard assessment: 12th volume. ASTM STP 1027, American Society for Testing and Materials, Philadelphia, PA.

Waltz, R.D. and Mccafferty, W.P. 1987a. Systematics of *Pseudocloeon, Acentrella, Baetiella* and *Liebebiella*, new genus (Ephemeroptera: Baetidae). J. N.Y. Entomol. Soc. 95: 553-568.

Waltz, R.D. and Mccafferty, W.P. 1987b. Revision of the genus *Cloeodes* Traver (Ephemeroptera: Baetidae). Ann. Entomol. Soc. Amer. 80: 191-207.

Waltz, R.D. and Mccafferty, W.P. 1987c. Generic revision of *Cloeodes* and description of 2 new genera (Ephemeroptera, Baetidae). Proc. Entomol. Soc. Wash. 89: 177-184.

Waltz, R.D., Mccafferty, W.P. and Kennedy, J.H. 1985. *Barbaetis*: A new genus of eastern North America. Great Lakes Entomol. 18: 161-166.

Warwick, R.M. 1988. The level of taxonomic discrimination required to detect pollution effects on marine benthic communities. Mar. Pollut. Bull. 19: 259-268.

Warwick, R.M. 1993. Environmental studies on marine communities: Pragmatical considerations. Aust. J. Ecol. 18: 63-80.

Warwick, R.M. and Clarke, K.R. 1991. A comparison of some methods for analysing changes in benthic community structure. J. Mar. Biol. Assoc. U.K. 71: 225-244.

Warwick, R.M. and Clarke, K.R. 1993. Comparing the severity of disturbance: A meta-analysis of marine macrobenthic community data. Mar. Ecol. Prog. Series 92: 221-231.

Waterhouse, J.C. and Farrell, M.P. 1985. Identifying pollution related changes in chironomid communities as a function of taxonomic rank. Can. J. Fish. Aquat. Sci. 42: 406-413.

Weber, C.I. (editor) 1973. Biological field and laboratory methods for measuring the quality of surface waters and effluents. Office of Water, United States Environmental Protection Agency, Cincinnati, OH. EPA/670/4-73/001.

Whitehurst, I.T. and Lindsey, B.I. 1990. The impact of organic enrichment on the benthic macroinvertebrate communities of a lowland river. Water Research 24: 625-630.

Wiggins, G.B. and Mackay, R.J. 1978. Some relationships between systematics and trophic ecology in nearctic aquatic insects, with special reference to Trichoptera. Ecology 59: 1211-1220.

Wilhelm, F.M. and Hiebert, J.A. 1996. A simple and inexpensive method to separate live and preserved benthos from sediments. J. Freshwat. Ecol. 11: 119-121.

Williams, C.J. 1985. A comparison of net and pump sampling methods in the study of chironomid larval drift. Hydrobiologia 124: 243-250.

Williams, D.D. and Williams, N.E. 1974. A counter-staining technique for use in sorting benthic samples. Limnol. Oceanogr. 19: 152-154.

Williams, N.E. and Hynes, H.B.N. 1973. Microdistribution and feeding of net-spinning caddisflies (Trichoptera) of a Canadian stream. Oikos 24: 73-84.

Winterbourn, M.J. 1985. Sampling stream invertebrates. In: Pridmore, R.D. and Cooper, A.B. (editors), Biological monitoring in freshwaters: Proceedings of a seminar, Hamilton, November 21-23, 1984. Part 2. Water and Soil Miscellaneous Pub. No. 83, National Water and Soil Conservation Authority, Wellington, New Zealand. p. 241-258. (Cited in Resh and McElravy 1993)

Wolcott, L.T., Griffith, M.B., Perry, S.A. and Perry, W.B. 1992. Evaluation of a pump core sampler in second order streams. J. Freshwater Ecol. 7(4): 419-424.

Wright, J.F., Moss, D., Armitage, P. and Furse, M.T. 1984. A preliminary classification of running-water sites in Great Britain based upon macro-invertebrate species and the prediction of community type using environmental data. Freshwat. Biol. 14: 221-256.

Wright, J.F., Armitage, P.D., Furse, M.T. and Moss, D. 1988. A new approach to the biological surveillance of river quality using macroinvertebrates. Int. Verein. Theor. Limnol. Verh. 23: 1548-1552.

Wright, I.A., Chessman, B.C., Fairweather, P.G., and Benson, L.J. 1995. Measuring the impact of sewage effluent on the macroinvertebrate community of an upland stream: The effect of different levels of taxonomic resolution and quantification. Aust. J. Ecol. 20: 142-149.

Wrona, F.J., Culp, J.M. and Davies, R.W. 1982. Macroinvertebrate subsampling: A simplified apparatus and approach. Can. J. Fish. Aquat. Sci. 39: 1051-1053.